Journal of
Experimental Psychology

VOL. 37, NO. 4 AUGUST, 1947

THE EFFECT OF PUNISHMENT ON DISCRIMINATION LEARNING IN A NON-CORRECTION SITUATION *

BY GEORGE J. WISCHNER [1]

State University of Iowa

I. INTRODUCTION

In recent years there have been reported a number of investigations concerned with a re-examination of the function of punishment in the acquisition and elimination of response tendencies. The results of these studies have been offered as support for a reformulation of that portion of the old law of effect pertinent to the action of annoyers.[2]

Muenzinger and his co-workers (7, 8, 9, 10, 12, 13,14) have been concerned especially with the function of electric shock in the discrimination learning situation. Their general program has been to vary systematically the time and locus of punishment and to compare the effects of electric shock with other punishing media.

Muenzinger's (7) first study stemmed from a methodological consideration. He questioned the adequacy of earlier studies such as those of Hoge and Stocking (3) when they are made the support for the generalization that punishment tends to weaken or inhibit the punished response.

> It is the so-called 'wrong' response that has always been punished in situations in which the conditions of placing the reward favored the performance of some alternative, the so-

* Read in part at the American Psychological Association meetings, Sept., 1946.

[1] This paper is based on a thesis submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the M.A. degree. The writer is indebted to Professor Kenneth W. Spence for suggesting the problem and for his constructive criticism. He is also grateful to Mr. Paul E. Griffith for his technical assistance.

[2] The formulation made by Thorndike (15) in 1911 is referred to here. More recently, Thorndike (16, 17) has restated that section of the law of effect dealing with the action of annoyers. Annoying after-effects are no longer viewed as dynamically opposites of satisfiers but as specialized in their action. The effect of annoyers seems to depend on "what they cause the animal to be or do." When punishment does weaken a connection, its effect is probably indirect in that the weakening is the result of the strengthening of some competing tendency or action.

> called 'right' response. In order to be sure that the function of punishment is really that of the inhibition of the punished response, we ought to punish the response favored by the reward, the 'right' response. In such a case we would expect according to the law of effect that a conflicting tendency be set up in a response that is both rewarded and punished, that is, a facilitating and inhibiting tendency which would manifest itself in a slowing down of the course of learning (7, p. 267).

Accordingly an experiment was devised to introduce the factor of shock for correct responses. The results indicate that shock for correct responses makes for as efficient learning as does shock for wrong responses. There were no significant differences between a shock-right group and a shock-wrong group in the learning by white rats of a black-white discrimination habit in a T-shaped apparatus, although the latter proved to be somewhat more efficient than the former. Both shock groups were significantly superior in learning efficiency to a group in which no shock was employed.

These data were offered by Muenzinger as a contradiction of that part of the 1911 statement of the law of effect dealing with the after-effects of annoyers and as support for the interpretation that the function of electric shock in discrimination learning is a general one. Shock does not weaken or inhibit the specific response that is shocked, concluded Muenzinger, but rather it makes the animal more sensitive to the cues to be discriminated.

This interpretation was later modified on the basis of data obtained in a subsequent experiment (**8**) in this series of studies. Although all shock groups were again significantly more efficient in learning than a no-shock group, the differences between the shock-right and shock-wrong groups were now statistically significant in favor of the shock-wrong group. It was proposed by Muenzinger that both general and specific mechanisms were operating in the use of shock.

Drew (**1**) and Fairlie (**2**) have reported studies concerned with this problem. Drew interprets his results as supporting the view that the function of shock is a general one. No significant differences in learning efficiency were found among the different shock groups; all shock groups were superior to a control no-shock group. However, as reported in the Muenzinger studies, the shock-wrong animals were better than the shock-right animals.

Fairlie applied shock to one group of animals just as the rat was about to enter the incorrect alley; a second group was shocked just as the animal was about to run into the correct alley. His results were quite different from those obtained by Muenzinger and Drew. Learning was markedly retarded in both shock groups, none of the shock-right animals reaching the criterion, and only three of the shock-wrong animals attaining the criterion. Again it is interesting to note that the shock-wrong group was superior to the shock-right group,

the former making an average of 38.2 errors, the latter an average of 50.2 errors.

II. Statement of the Present Problem

In the Muenzinger experiments, all animals, irrespective of the conditions under which they were run, eventually approached the correct food box and were rewarded on every trial before the trial was terminated. That is, they were permitted to correct in the event of a wrong response. Hull and Spence (4), in an analysis based on conditioned response principles, have shown that differential behavioral phenomena may obtain in a simple T-maze trial-and-error situation, depending on whether a correction or non-correction procedure is employed.

TABLE I

An Analysis of the Correction and Non-Correction Procedures Showing the Consequents of Correct and Incorrect Responses

	Non-correction		Correction	
	+	−	+	−
No-shock	F	o	F	o, F
Shock-wrong	F	S	F	S, o, F
Shock-right	S, F	o	S, F	o, S, F

(+)—correct response; (−)—incorrect response; F—food; S—shock; O—'nothing.'

Table I presents the differences between the two procedures in terms of the consequents of a correct or incorrect response for the three conditions employed in the Muenzinger experiments. It is readily apparent that with both methods identical consequents result for each condition when the animal makes a correct choice, with the obvious exception in the case of the shock-right group. The contrast between the two methods is most readily seen in the instance of a wrong choice. With the non-correction procedure, there is a single consequent—either 'nothing' or a shock. When correction is permitted, the animal must necessarily receive some combination of incentives. It is interesting that the combination of consequents for an incorrect choice is the same for both shock groups, although the consequents follow in different sequence and occur in different parts of the apparatus.[3]

For a systematic approach to an understanding of the effect of punishment on discrimination learning it appears that the non-correction procedure has the following advantages over the correction

[3] The situation was further complicated in the first Muenzinger study. The shock-wrong animals were shocked both on entering and on leaving the incorrect alley.

procedure: (1) The animal can respond to (approach) only one stimulus cue in a given trial. (2) Only one consequent follows a choice-food, shock or 'nothing.' The obvious exception is in the case of a correct choice by the shock-right animals. (3) The time interval between a response and its consequent is better controlled and is relatively the same for all conditions. (4) All animals must make essentially the same response to the shock. The animal must continue over the grid in a forward direction toward the stimulus cue. (5) The total distance traveled by an animal in the course of a run following a choice is constant for all animals and for all trials. Since no withdrawal is possible, the animal, having entered an alley, must traverse its full length into the goal box where it may or may not be rewarded.

An examination of the literature reveals few studies in discrimination learning which have not been complicated by permitting correction. This is stated explicitly or may be inferred from description of the animals' behavior. Drew (1) remarks that the animals in his shock-on-right-door group, on receiving the shock, would withdraw, retrace to the incorrect door, and spend some time attempting to open it before returning to the correct shock door and finally going through it.

Warden and Aylesworth (19) attempted to eliminate correction but did not prevent return to the reaction compartment after an incorrect response. Moreover, some of the animals ran around to the correct alley before the experimenter could remove them from the apparatus.

Wood (20), in a discrimination learning study concerned with delayed reward and punishment, employed a non-correction procedure and criticized the methodology of earlier discrimination learning studies. Either the animal has been allowed to escape to the reaction compartment, or has received food in addition to shock for a wrong response on any particular trial, or, as in the case of the Hoge and Stocking experiment, it is not clear whether correction was or was not controlled. In his presidential address, Tolman (18) very pointedly indicated his awareness of the general issue—correction versus non-correction.

> A second point about which we are still surprisingly ignorant is that we do not yet know the importance of the rat's being permitted, or not permitted, to return out of the wrong choice. In some experiments, when the animal takes the wrong alley, he passes through a one-way gate and is started over again. In others, he is allowed to treat it as a blind and back out. But, so far as I know, there has been no carefully controlled comparison between these two procedures (18, p. 5).

The present experiment was designed to study the effect of electric shock for correct and incorrect responses on the acquisition of a

visual discrimination habit when the animal is prevented from correcting or withdrawing following a choice, whether correct or incorrect.

III. Experimental Procedure

A. Apparatus

The apparatus, a floor plan of which is shown in Fig. 1, was a modified Yerkes-Watson discrimination box. The contact doors represented an important feature. They prevented with-

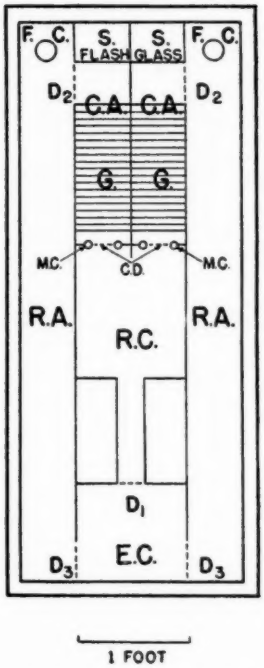

Fig. 1. Ground plan of apparatus. E. C.—Entrance Compartment; R. C.—Reaction Compartment; C. A.—Choice Alleys; G—Copper Grid; S—Stimulus Compartments; F. C.—Food Cups; R. A.—Return Alleys; D_1, D_2, D_3—Doors; M. C.—Mercury Cups; C. D.—Contact Doors.

drawal from an alley and so correction of a wrong response, and they formed part of the circuit to the grill. Each door was weighted by attaching pieces of galvanized tin of the same dimensions as the door. The door could be quickly lowered by depressing a spring clip, thus releasing the fishline holding the door in the 'up' position. Pieces of sponge rubber cemented under the ends prevented the door from hitting against the floor of the apparatus.

Soldered to each contact door, on the metal side, were two stiff pieces of copper wire which dipped into mercury cups when the contact doors were released. A single pole, double-throw knife switch was connected in the circuit in such a way that its position determined which door completed the circuit.

The grill was a slab of bakelite about which was wound number 20 copper wire. Two separate lengths of wire were wound so that they made alternate turns ¼ in. apart. The shock stimulus was a direct current of 20, 25, or 30 microamperes. With the type of electrodes employed, this represented approximately twice the threshold for rats of the ages used.[4] It has been demonstrated that the rat's bodily resistance level averages about 300,000 ohms (11). With the circuit used, any resistance changes between 0 and 500,000 ohms produced no measurable changes in the current. The source of the current was a self biased type 6c6 pentode vacuum tube provided with a plate voltage of 650 volts. The magnitude of the current could be adjusted by means of a variable bias voltage obtained from a battery and its associated potentiometer.

Other notable features of the apparatus were the return alleys which eliminated trial-to-trial handling of the animal by the *E*, and the proximity of the goal and stimulus compartments.

At the end of each stimulus compartment was a 40-watt bulb, the light from which was diffused by means of flash glass placed in front of it. Either bulb could be illuminated by manipulating a double-throw switch.

All doors were of plywood, excepting those leading from the choice alleys into the goal compartment, which were made of transparent celluloid material. The doors, lined with felt, were manipulated in a vertical fashion by means of fishline and were practically noiseless. D_3 and D_2 were counterbalanced in such fashion that when the former was brought into a 'down' position, the latter was raised.

Food was supplied by means of a drawer-like arrangement. The 'drawer,' a small block of wood in which was drilled a hole, could be slipped in under the apparatus so that the food coincided with an opening drilled in the floor of the apparatus itself. From a position behind a one-way screen situated back of the stimulus end of the apparatus, the *E* controlled all light and shock switches, the food cups, and the strings operating the various doors.

B. Subjects

Thirty white rats, 17 males and 13 females, from the colony maintained at the State University of Iowa, were used in the experiment. They were distributed in approximately the same proportion with reference to sex among the three groups of 10 rats each employed in the study. The animals were fed approximately a half hour after they had completed the day's work. They ranged between the ages of 64 days and 91 days at the beginning of the experiment.

C. Procedure

Preliminary training.—This covered a period of three days. Day 1 served as adjustment to the apparatus. All doors were removed, animals were introduced into the apparatus in pairs and were allowed to explore freely for 20–30 min. They were permitted to find food (Purina pellets) in both food cups several times. Days 2 and 3 were for training in the return habit. All doors were in place, D_3 and D_1 being down and D_2 raised. One contact door was down. An animal was placed in the entrance compartment and D_1 raised. As the rat advanced from the reaction compartment to the food compartment through the open choice alley, D_1, the contact door and then D_2 were lowered. When the animal had found and eaten the food it would soon proceed fairly quickly of its own accord along the return alley to the entrance compartment where it was in position for the next trial. Each *S* was given 10 such trials per day for two days with a food reward at the end of each trial. The position of the open alley was varied according to the prearranged orders: RLLRRLRLLR, LRRLLRLRRL. These orders were also used for varying the position of the positive stimulus during regular training, one being used on odd days and the other on even days.

Classification series.—The procedure on this day was the same as that employed during regular training except that shock was not introduced. Both contact doors were up at the beginning of each trial and the cue stimuli were presented, the position of the positive cue being varied in accordance with the orders indicated above. The animals received food if they responded to the lighted alley (positive stimulus) and nothing if they responded to the dark alley (negative stimulus). On the basis of their scores on this day (percent responses to the lighted alley), the animals were divided into three groups as follows: shock-right, shock-wrong, and no-

[4] Rough threshold determinations were obtained with 15 animals. In all instances, a reaction was elicited at from 10 to slightly more than 15 microamperes.

shock. An attempt was made to allocate animals to the three groups so that the mean correct responses on Day 4 were approximately the same for all groups. The means and standard errors of the means for the three groups in the order mentioned were 60 ± 4.2, 60 ± 4.7, and 57 ± 4.5.[5]

Method in learning problem.—The regular learning series was begun on Day 5. The shock-right animals received both shock and food if they responded to the lighted alley and 'nothing' in the dark alley; the shock-wrong animals were shocked in the dark alley and received food in the lighted alley; the no-shock animals were rewarded for a response to the light and frustrated for responding to the dark. An animal was considered to have made a choice and the contact door was released when its four feet were on the grill and its nose reached an arbitrarily designated point in the choice alley.

For the first six days of regular training, the shock was kept at 20 microamperes. It was increased to 25 microamperes on the seventh day and to 30 microamperes on the eighth day, at which intensity it was maintained until the animal attained the learning criterion.

All animals were given five trials a day for the first six days and 10 trials daily thereafter until they attained the criterion of 90 percent correct in 20 trials, the last 10 of which were all correct. There was a 20- to 30-sec. interval between trials.

If an animal failed to make a choice within a five-min. period, it was removed from the apparatus to a carrying cage and the trial was counted and recorded as NR (no reaction). A further opportunity to work was given after another animal had been run. If an animal failed to respond on successive trials given in this manner NR's were recorded until the number of trials equaled that of the usual number given.[6]

IV. Results

The principal results are expressed in terms of the average number of errors for the total learning period and the average number of trials required to attain the criterion of mastery.[7] These data for the three groups are presented in Table II. The significance of the differences between groups for both measures is given in Table III in terms of *t*-values.

TABLE II

Mean Trial and Error Scores

Groups	Trials	σM	Errors	σM
Shock-right	159.0	14.3	71.3	10.7
Shock-wrong	104.0	6.0	24.2	3.3
No-shock	152.0	10.9	54.6	4.9

It is evident that with respect to both trials and errors, the shock-wrong group is superior to the other two groups. The differences

[5] This procedure suggests an experimental design involving matched groups. The correlation between performance on Day 4 and the measures of learning used proved to be negligible and consequently no correction for correlation was made in the calculation of the standard error of the mean differences.

[6] Very few NR trials were recorded, only two shock-right animals and three shock-wrong animals manifesting such responses. Almost all such responses occurred within the first 20–30 trials.

[7] Muenzinger and his students have presented their results in terms of errors for the first 100 trials. A comparable analysis was made for the present experiment, but since it showed the same general picture as the complete learning records the results of this analysis have not been included here.

TABLE III

SIGNIFICANCE OF GROUP DIFFERENCES †

	Trials		Errors	
	Difference	t	Difference	t
Shock-wrong vs. shock-right	55.0	3.5*	47.1	4.2*
Shock-wrong vs. no-shock	48.0	3.9*	30.4	5.2*
No-shock vs. shock-right	7.0	0.4	16.7	1.4

* Significant at better than the one percent level of confidence.
† The more efficient group of a pair is given first.

between this group and each of the other groups are significant at better than the one percent level of confidence.[8] The shock-right group is the least efficient in learning, although the differences between this group and the no-shock group are not significant (t = 1.4 for errors and 0.4 for trials).

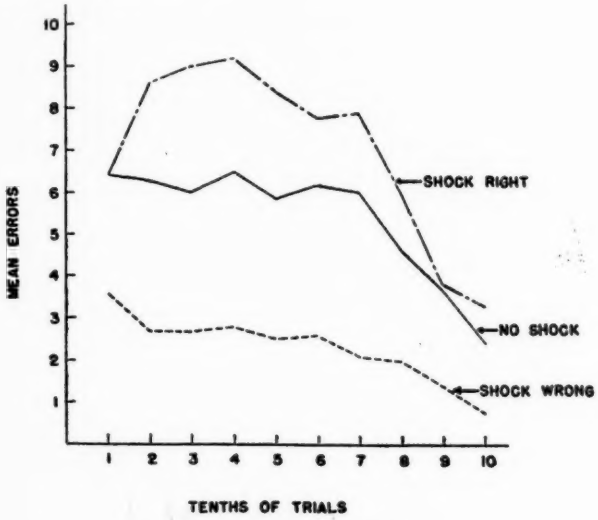

FIG. 2. Vincent error curves

The results are given graphically in Fig. 2 in the form of Vincent error curves for each group.[9] These curves show even more strik-

[8] A t-value of 2.89 falls at the one percent level of confidence with the number of degrees of freedom in this study.

[9] The procedure employed for constructing these curves was based on Hunter's (5) discussion.

ingly the relative learning efficiency of the three groups. That the shock-right group is inferior even to the no-shock group, particularly in the early stages of learning, is readily apparent. Of special interest is the initial rise in errors exhibited by this group. An examination of the individual records reveals that eight of the 10 *S*s in this group show this initial increase in incorrect responses, some of them choosing the dark alley 100 percent of the time. Of equal interest is the relatively rapid decline in the latter part of the curve when the animals finally begin to learn.

V. Discussion of Results

In the Muenzinger and Drew experiments the shock-after-choice groups were consistently superior to the no-shock groups and the shock-wrong condition resulted in faster learning than the shock-right condition. The results of the present study agree with the findings of these earlier investigations so far as the direction of the differences between the two shock groups is concerned. However, the magnitude of the differences was much greater in the present experiment, a result which, it will be recalled, was largely due to the very retarded learning of the shock-right animals. In fact these animals learned even more slowly than the no-shock group.

The finding that the learning of the shock-right group was accelerated as compared with the no-shock group in Muenzinger's experiments led him to deny that portion of Thorndike's (**15**) old law of effect dealing with the function of punishment. As this older formulation would imply, the response of approaching the positive stimulus would be weakened in the shock-right group with the consequence that learning would be slowed down. The conflicting evidence of the present experiment could be taken as supporting the old law of effect. However, as has been noted earlier, Thorndike (**16**, **17**) has himself abandoned this section of his law and has offered a different interpretation of the action of punishment. Our own explanation of these experimental data incorporates the principle expressed in this newer formulation.

Before we proceed to our interpretation, however, attention should be directed to the fact that the results of the present study do not support Muenzinger's generalization that shock after choice always has a facilitating effect. In the present situation the learning of the shock-right animals was retarded rather than facilitated.

Muenzinger has offered both a general mechanism and also specific mechanisms as operating in the function of punishment. His early findings led him to emphasize what he referred to as a general factor, one which increased the likelihood of reception of the relevant

stimulus cues. The results of a subsequent experiment (**8**) led him to supplement this general factor by postulating specific mechanisms, one operating in the case of the shock-right and one in the case of the shock-wrong animals. These specific mechanisms were introduced to account for the differences in the effects of shock for right and wrong responses. However, Muenzinger is rather vague as to the nature of these so-called specific mechanisms and alludes merely to certain differences in the behavior of the shock-right and shock-wrong animals in the negative, dark alley. He states that the latter learned very quickly to turn about before they advanced very far into the wrong alley whereas the former continued to the end of this alley.[10]

The present writer is inclined to agree that shock has a general function in the sense that it slows down the running behavior of the animal, producing a pause which increases the duration of stimulation, and tends to decrease responses made in disregard of the relevant stimulus cues. But our results indicate that shock also has an important specific function, the nature of which may be formulated more precisely than the vague specific mechanisms proposed by Muenzinger.

The findings of the present study favor the hypothesis that this highly important function of shock is specific with respect to the cues associated with punishment. The animal very quickly tends to acquire avoidance responses (anticipatory withdrawal) to the cues which are present during the shock experience.[11] The strongest evidence for this specific effect is to be found in the relatively marked retardation of our shock-right group and especially in the error curve for this group. From the early drop in the error curve of the shock-wrong animals, one may infer that the same principle was operating for this group. The shock-right animals learned very quickly to avoid the lighted alley; the shock-wrong group learned very quickly to avoid the dark alley.

[10] It should be noted that the conditions of the present experiment did not permit differences in degree of penetration into the wrong alley following a choice. Having made a final response, correct or incorrect, the animal was forced to continue the full length of the choice alley into the food compartment.

[11] There is interesting qualitative evidence for the 'anticipatory' character of this withdrawal. An animal could not receive a shock until after it made a choice and the contact door was released. Even though the shock was not experienced, the shock-right animals in the early stages of learning would look at the lighted alley just once, then turn and run down the dark alley. Further indications of the 'anticipatory' nature of this withdrawal are found in the following behavior exhibited by practically all the shock animals. As the animal entered the reaction compartment and faced the stimuli it often made a quick jerky backward movement. This was similar to the initial reaction of an animal experiencing its first shocks when it would back up against the contact door, attempting to withdraw, and then finally run across the grid. Such behavior might be exhibited when an animal chanced to touch the uncharged grill as it vacillated between the two alleys before making a choice.

Applying this interpretation of the function of punishment to the present experiment, it is not difficult to understand the superior performance of the group shocked for incorrect responses. The specific effect of shock favored the learning of the problem by these animals because it tended to make them avoid the negative stimulus with resulting fewer errors. At the same time the approach response to the positive stimulus was strengthened early in the training series through the process of frequent reinforcement (food reward). In the case of the shock-right animals the situation is more complex. In this group, although the response of approaching the light was reinforced by food, the response of withdrawing from this stimulus was reinforced by escape from shock. There ensued a competition between approaching light and obtaining food, and withdrawing from this cue and escaping shock. In addition there was also present the competing tendency to choose the dark alternative alley, as on some occasions this alley, insofar as its spatial component is concerned, i.e., right or left, had led to food. Of equal significance is the fact that the dark alley led to escape from shock and termination of the trial.

There are indications in the Muenzinger and Drew reports which appear to support the assumption that this specific effect of shock, which manifests itself in the acquisition of avoidance behavior to cues associated with it, was also operating to some extent in their situations. Drew reports that for the first 20 trials his shock-on-right-door group made more errors than his no-shock group. Also relevant in this connection are Muenzinger's behavioral observations to the effect that his shock-right animals tended to run to the end of the alley in which they were not shocked while his shock-wrong animals learned very quickly to turn around in the alley in which they were shocked.

Assuming that such similar specific mechanisms were operating in these experiments, how is one to interpret the relative accelerating effect of shock after choice for correct responses in these studies and the apparent absence of any accelerating effect for the corresponding group in the present study?

It is believed that the differences may be, in part, a function of the methodology employed, namely, correction or non-correction. In the correction procedure it is only the 'correct' response, with its consequents of shock plus food, that terminates each trial. Going to the dark alley or door does not terminate the trial. The animal must on every occasion ultimately cross the grill in the lighted alley or proceed through the lighted door and thus receives food reinforcement in addition to the shock. On every trial, then, there may be a weakening of the approach response to dark (frustration) and a strengthening of the approach response to light (food reinforcement). The methodology

of non-correction, on the other hand, affords an alternative response for the shock-right animal, that to the dark alley, which may complete the trial with what appears to be the single consequent of frustration. This response to the dark alley may be rewarding, however, the reinforcing mechanism being the tension reduction resulting from the escape from the noxious stimulus and termination of the trial. It would appear that the conditions of correction, which 'force' the animal to approach the light and thus guarantee reinforcement of this response on each trial, make for a more rapid shift in the relative strengths of the light and dark stimuli to evoke the approach response to light, than do the conditions of non-correction.

If the function of shock is to establish avoidance responses to cues associated with the shock experience, how do the shock-right animals finally come to learn, that is, to choose the lighted alley consistently in spite of the shock? An answer to this question is suggested by Mowrer's (**6**) observations that rats, when subject to a shock, tend to adopt responses which will lessen the effective intensity of the shock. His animals learned that by remaining quietly on their hind legs and lifting their forepaws off the grill, they received relatively little shock. Mowrer has pointed out that with a circuit with characteristics similar to that used in the present experiment, this mode of adjustment would have increased rather than decreased the subjective intensity of the shock.[12] Therefore, even more significant is the manner in which the animals in this study attempted to adapt to the shock stimulus. The shock-right animals very characteristically would get set before the lighted alley, hop on the grill landing on all four paws, and then hop off, clearing the grill in this fashion with a single contact. It is highly probable that this acquired mode of adjustment lessened the pain stimulation and decreased the potency of the shock stimulus as a determining factor in the situation.

These acquired adjustive acts may further explain the relatively quicker learning of the shock-right animals in the correction experiments. Two factors have already been suggested, namely, the frustration of the tendency to approach the negative (dark) stimulus which led to neither food, escape from shock, nor termination of the trial, and the reinforcement of the approach response to the lighted alley on every trial. These animals had to experience the shock on each trial and these successive experiences with the shock from the beginning of training may have led to more rapid acquisition of the modes of adjustment which tended to decrease the pain.

[12] Because of the relatively constant current which was maintained in spite of very large changes in resistance, the smaller the area of contact with the electrodes, the greater would be the density of the current at the points of contact, and the greater, therefore, would be the effective intensity of the shock.

VI. Summary

1. This experiment was designed to study the effect of electric shock on the acquisition of a visual discrimination habit when the *S* is prevented from correcting or withdrawing following a choice.

2. Three groups of white rats were trained to go to the lighted alley in a modified Yerkes-Watson discrimination box. A shock-right group received both shock and food for a response to the lighted alley and 'nothing' for a response to the dark alley. A shock-wrong group was given food for responding to the lighted alley and a shock for responding to the dark alley. A third no-shock group received food for a correct choice and 'nothing' for an incorrect choice. The apparatus was designed so that, having made a choice, the animal could not withdraw but had to continue in a forward direction toward the stimulus cue.

3. The shock-wrong group was superior to the other two groups in terms of both trials and errors. The mean trial and error differences between the shock-wrong group on the one hand, and the remaining two groups on the other, were statistically significant. The shock-right animals were least efficient although the mean trial and error differences between this group and the no-shock group were not statistically significant. The error curve for the shock-right *S*s is characterized by an increase in errors in the early stages of learning and a relatively rapid decline in the later trials when the animals finally begin to learn.

4. It is concluded that shock has a specific function with respect to the response that is punished. The animal very quickly builds up avoidance responses to the cues associated with shock.

5. These results are seemingly in conflict with the findings of other studies that shock administered anywhere after choice produces a general accelerating effect.

(Manuscript received September 20, 1946)

References

1. Drew, G. C. The function of punishment in learning. *J. genet. Psychol.*, 1938, **52**, 257–266.
2. Fairlie, C. W., Jr. The effect of shock at the 'moment of choice' on the formation of a visual discrimination habit. *J. exp. Psychol.*, 1937, **21**, 662–669.
3. Hoge, M. A., & Stocking, R. J. A note on the relative value of punishment and reward as motives. *J. Anim. Behav.*, 1912, **2**, 43–50.
4. Hull, C. L., & Spence, K. W. Correction vs. non-correction method of trial-and-error learning in rats. *J. comp. Psychol.*, 1938, **25**, 127–145.
5. Hunter, W. S. Learning: IV. Experimental studies of learning. In: C. Murchison (Ed.) *A handbook of general experimental psychology*. Worcester, Mass.: Clark University Press, 1934.
6. Mowrer, O. H. An experimental analogue of regression with incidental observations on reaction formation. *J. abnorm. soc. Psychol.*, 1940, **35**, 56–87.
7. Muenzinger, K. F. Motivation in learning. I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 267–277.

8. MUENZINGER, K. F., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning. VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.
9. MUENZINGER, K. F., & FLETCHER, F. M. Motivation in learning. VI. Escape from electric shock compared with hunger-food tension in the visual discrimination habit. *J. comp. Psychol.*, 1936, **22**, 79–91.
10. MUENZINGER, K. F., & FLETCHER, F. M. Motivation in learning. VII. The effect of an enforced delay at the point of choice in the visual discrimination habit. *J. comp. Psychol.*, 1937, **23**, 383–392.
11. MUENZINGER, K. F., & MIZE, R. H. The sensitivity of the white rat to electric shock: Threshold and skin resistance. *J. comp. Psychol.*, 1933, **15**, 139–148.
12. MUENZINGER, K. F., & NEWCOMB, HELEN. Motivation in learning. III. A bell signal compared with electric shock for right and wrong responses in the visual discrimination habit. *J. comp. Psychol.*, 1935, **20**, 85–93.
13. MUENZINGER, K. F., & NEWCOMB, HELEN. Motivation in learning. V. The relative effect of jumping a gap and crossing an electric grid in a visual discrimination habit. *J. comp. Psychol.*, 1936, **21**, 95–104.
14. MUENZINGER, K. F., & WOOD, ALDA. Motivation in learning. IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J. comp. Psychol.*, 1935, **20**, 95–106.
15. THORNDIKE, E. L. *Animal intelligence; experimental studies.* New York: Macmillan Co., 1911.
16. THORNDIKE, E. L. *The fundamentals of learning.* New York: Teachers College, 1932.
17. THORNDIKE, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton Century, 1935.
18. TOLMAN, E. C. Determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
19. WARDEN, C. J., & AYLESWORTH, M. The relative value of reward and punishment in the formation of a visual discrimination habit in the white rat. *J. comp. Psychol.*, 1927, **7**, 117–127.
20. WOOD, A. B. A comparison of delayed reward and delayed punishment in the formation of a brightness discrimination habit in the chick. *Arch. Psychol.*, New York, 1933, **24**, No. 157.

STUDIES IN SPATIAL LEARNING. V. RESPONSE LEARNING VS. PLACE LEARNING BY THE NON-CORRECTION METHOD

BY E. C. TOLMAN, B. F. RITCHIE, AND D. KALISH[1]

University of California

Introduction

In the second article (2) of this series it was reported that rats trained by the correction method[2] were relatively unable to learn to make the same response, say a right turn, to two different places. Of the eight rats trained to run in half of the trials from S_1 to F_1 and in half of the trials from S_2 to F_2 on a simple elevated T-maze, the starting-leg of which could be attached to either side of the arms of the T (cf. Fig. 1), only three were able to learn the problem. A recent investigation by Kalish (1) of the non-correction method suggests, however, that different results might have been obtained if the non-correction method had been used instead of the correction method. The experiment to be reported in the present paper was therefore initially designed to determine if rats could learn relatively easily to run from S_1 to F_1 and from S_2 to F_2 if trained by the non-correction method. Secondly, we wanted to know if they would still learn less readily than a place-learning group[3] also run by the non-correction method. Thirdly, we wanted to know, if rats trained by the non-correction method can develop the ability to make a single response, e.g., a right turn, to get to two different places, what generality will this response have? For instance, if the rats learn to turn right when started at either S_1 or S_2 when the arms of the T run east and west, will they be able to maintain this response when the apparatus is rotated 90 degrees around the choice point, that is, when the arms of the T run north and south? Finally, a fourth question arose, namely, how would a place-learning group, run as a control, behave when it also was rotated through 90 degrees?

[1] The cost of this investigation was met in part by grants to the Department of Psychology from the Research Board of the University of California.

[2] 'Correction method' refers to a training procedure in which an animal, if he makes a wrong choice and runs to the end of the blind, is allowed to turn around and run to the goal. In contrast, the 'non-correction method' is one in which if an animal makes the wrong choice and runs to the end of the blind, he is immediately removed from the maze and placed again on the starting-leg. This procedure is continued in each trial until the animal makes the correct choice.

[3] In the second article of this series (2) a 'place-learning group' was defined as one in which food was found always at the same geographical end of the maze, say at F_1, whichever side the starting-leg was put.

To sum up, there are four questions which the present experiment sought to answer: (1) can rats, if trained by the non-correction method, learn relatively easily to make the same response to get to two different places; (2) how will this group's learning rate compare

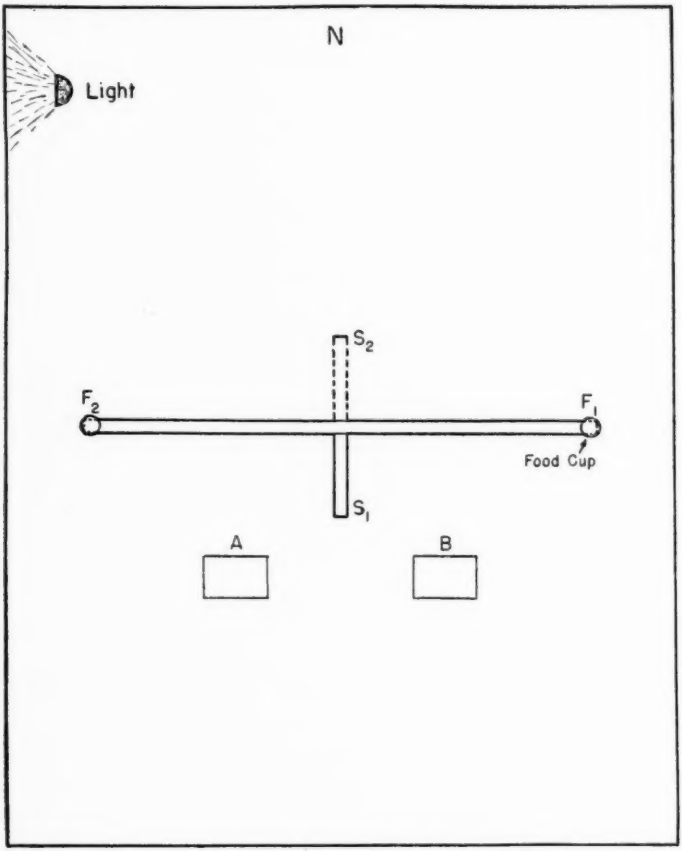

FIG. 1. Floor plan of maze before rotation

with that of a place-learning group also run by the non-correction method; (3) how, after they have learned in the original set-up, will the response-learning group behave on a critical trial given after the maze has been rotated through 90 degrees; and finally, (4) how will the place-learning group behave on a critical trial if, after they have learned, the maze be rotated through 90 degrees?

Animals

Two groups, one containing seven and the other containing eight male M × M pigmented rats from the Tryon 'bright' and 'dull' stocks, were used in this experiment. They were all between 90 and 150 days old when the experiment was performed. None of the rats was maze-wise.

Apparatus

The apparatus consisted of a simple elevated T-maze with a detachable starting-leg. The arms of the T were 1.5 meters long; the starting-leg was .5 meter long (cf. Fig. 1). A paper cup, five cm. in diameter, was attached to the end of each arm. Food was placed in the cup which was at the end of the correct arm. The maze was located in a room 5 meters long, 4 meters wide, and 3.5 meters high. The maze was placed so that the choice point was exactly in the center of this room. The only source of light was a desk lamp placed two meters from the floor in the north-west corner of the room. The light shone onto the wall and reflected back across the room, giving definite asymmetrical lighting (cf. Fig. 1). The only other objects in the room (the rats lived in a separate room) were two small boxes, A and B, which served to support the cage of animals being tested at a given time.

Two groups of rats were run: (1) an experimental group called 'the response-learning group,' that is, a group that had to learn to make the same response to get to two different places; and (2) a control group called 'the place-learning group,' that is, a group that had to learn to make two different responses (a right turn and a left turn) to get to the same place.[4] Both groups were run by the non-correction method.

Preliminary training: Forty-eight hours before the preliminary training was begun the rats were placed on a 24-hour wet-food maintenance schedule, being fed at 10 in the evening. Preliminary training, which lasted three days, consisted of familiarizing the animals with handling, elevated paths, and eating from the paper cup attached to the end of a single elevated path.

Response-learning group.—The day following the end of the preliminary training the seven rats of this group were given four trials on a simple T-maze. (This was not the maze used in the experiment proper.) Food was at both ends of the maze. On each trial the maze was set in a different spatial direction. That is, on the first trial the animals when placed on the starting-leg faced north; on the second trial when placed on the starting-leg they faced east; etc. This test for an initial turn habit gave negative results; that is, none of the animals consistently turned to the right or to the left. Neither this test nor the preliminary training was given in the room in which the experiment proper was performed.

On the day following that on which the turn habit was tested the experiment proper was begun. In view of the negative results of the turn habit test the animals were divided arbitrarily into two subgroups. For the one subgroup a right turn was correct; for the other a left turn was correct. Four trials were given on each night. On the first trial the starting-leg was attached to the south side of the arms of the T (cf. Fig. 1); on the second and third trials it was attached to the north side, and on the fourth trial it was again attached to the south side of the T. The non-correction method was used throughout; that is, if an animal made the incorrect turn he was removed from the apparatus when he reached the end of the blind and was immediately placed back on the starting-leg. This procedure was continued until the animal made the correct turn. A trial was not counted complete until such a correct turn had been made. An animal was allowed to eat at the end of the correct path for 25 sec. before he was returned to the cage. If instead of eating for the full 25 sec., the rat, after having taken a bite of food, attempted to retrace, that is, attempted to return to the choice point, he was immediately picked up and returned to the cage.

The cage in which the rats were brought into the room where the experiment was performed was placed on either box A or box B, an empty cage being on the other box (cf. Fig. 1). These boxes, located a little behind and one to each side of the starting-leg, maintained a fixed relation to the starting-leg; that is, they were shifted to the other side of the room when the starting-leg was shifted to the other side. After an animal had completed his trial he was placed in the cage

[4] We already know from the second article of this series (2) that a place-learning group, that is, a group that has to learn to run to, say F_1, from either S_1 or S_2 (cf. Fig. 1), learns rapidly if trained by the correction method.

from which he had not been taken. After all the animals of a cage had completed their trial, they were returned to the home room while another cage of animals was run. *E* stood directly behind the starting-leg between boxes A and B.

Intra-maze cues were shifted in a random fashion by rotating the arms of the T around the choice point between trials.

The rats were trained by this procedure for 10 days. On the eleventh day, after two of the four trials had been given, the maze was rotated 90 degrees, that is, the arms of the maze now ran north and south instead of east and west, and the starting-leg was attached to the west side of the arms of the T. In this new position a 'critical trial' was run in the same manner as the preceding trials.

Place-learning group.—The day following the preliminary training the eight rats of this group were given four trials to determine initial position habits. These four trials were run in the home room with the arms of the T (not the T-maze used in the experiment proper) running east and west, with the lighting conditions similar to those under which the experiment proper was to be performed, and with the procedure identical to that which was to be used in the experiment proper except that food was at the end of both arms.

The day following that on which the position habit test was run the experiment proper was begun. In view of the results of the position habit test, the animals were divided into two subgroups. Those that had gone predominantly to the east were placed in a subgroup which found food at the end of the west arm; those that had gone predominantly to the west were placed in a subgroup which found food at the end of the east arm; and those animals that had manifested no preference were distributed arbitrarily between both subgroups. Each subgroup contained four animals.

The training procedure for this place-learning group was identical with that given to the response-learning group, except that the training, because of more rapid learning, was continued for only six days. On the seventh day, after two of the four trials had been given, the maze was rotated 90 degrees and the starting-leg attached to the west side of the arms of the T. With the maze in this new position a critical trial was run in a manner similar to that of the preceding trials. That is, the new correct goal locations for this critical trial were decided upon in terms of the relation of the ends of the maze to the light source, this latter being thought to be the only cue by means of which the animal could maintain his spatial orientation. The east and south sides of the room are farther from the light than the west and north sides. Hence, it was decided that the subgroup which had found food at the end of the east arm should now find food at the end of the south arm and that the subgroup which had found food at the end of the west arm should now find food at the end of the north arm.

Results

Response-learning group.—Of the seven rats in the response-learning group, three had reached a criterion of two consecutive errorless days on the sixth day of training (24 trials); three others had reached this criterion by the ninth day (36 trials); the other animal was still running at chance when the maze was rotated on the 11th day. No animal made an error after he had reached the criterion. During the last part of their training all animals ran unhesitatingly to the end of the arm they turned into. On the tenth day, that is, the day before the critical trial, the mean time per rat per trial to make a choice (i.e. the time at which an animal was placed on the starting-leg until the time his body had completely entered one of the arms) was four sec. VTE's had almost entirely disappeared by this day.

The performance on the critical trial given on the 11th day, that is, the trial following the rotation of the maze, was quite new. On this trial all the rats (including the one which had not yet learned the

problem) made the correct turn. The mean time per rat to make a choice was the same as it had been on the preceding day, namely, four sec. VTE's increased slightly; the average was one per rat instead of one for the group as it had been on the preceding day. But, and this is what was new, no rat after turning ran unhesitatingly to the food cup at the end of the arm he turned into. Each rat stopped running when he had covered approximately half of the distance from the choice point to the food. After this halt, all the rats except the one that had not learned the original problem crept slowly to the food cup, ate a small amount of food, and then ran toward the other end of the maze. The animal that had not learned the original problem turned and ran toward the other end of the maze.

Place group.—Of the eight rats run in the place-learning group, six had reached a criterion of two errorless days by the end of the fifth day (20 trials). Of the other two animals, one had an errorless performance on days both four and six, but made an error on the fifth day; the other animal made an error on both day four and day six, and two errors on the fifth day. On the sixth day, that is, the day before the critical trial, the mean time per rat per trial to make a choice was eight sec.; the animals averaged one VTE per trial on this day.

On the critical trial, given on the seventh day, six of the eight animals made an error, that is, ran to the wrong end of the maze. The mean time per rat to make the choice was seven sec., and there was an average of one VTE per animal. All of the animals ran unhesitatingly to the end of the arm they turned into.

Fig. 2 contains learning curves for the two groups for the training preceding the critical trial.

Discussion

To the initial question which we raised the results of the present experiment give an affirmative answer. Rats trained by non-correction method *can* learn relatively easily to make the same response to get to two different places. Although when trained by the correction method rats had great difficulty in learning this problem, in view of the analysis of the non-correction method given in another article by one of the present writers (1) this conclusion was to have been expected. Kalish there reported that the non-correction method, by preventing correction of erroneous responses (i.e., by preventing reinforcement of such responses as S_1–F_2–F_1 and S_2–F_1–F_2), broke up consistent place-going position habits. Most of the animals in the response-learning group reported in the second article of this series (2), when trained by the correction method, developed such east end or west end habits immediately, and in maintaining them

failed to learn the problem. Kalish in his analysis of the non-correction method also reported that both 'contrast by temporal contiguity,' that is, the close temporal contiguity between a final erroneous response and the succeeding correct response, and 'frequency,' that is, the necessary occurrence of the correct responses (i.e., direct runs from S_1 to F_1 and from S_2 to F_2) a fixed number of times each day, are also important factors in learning. All three of these by-products of the non-correction method apparently helped the response-learning group in the present experiment.

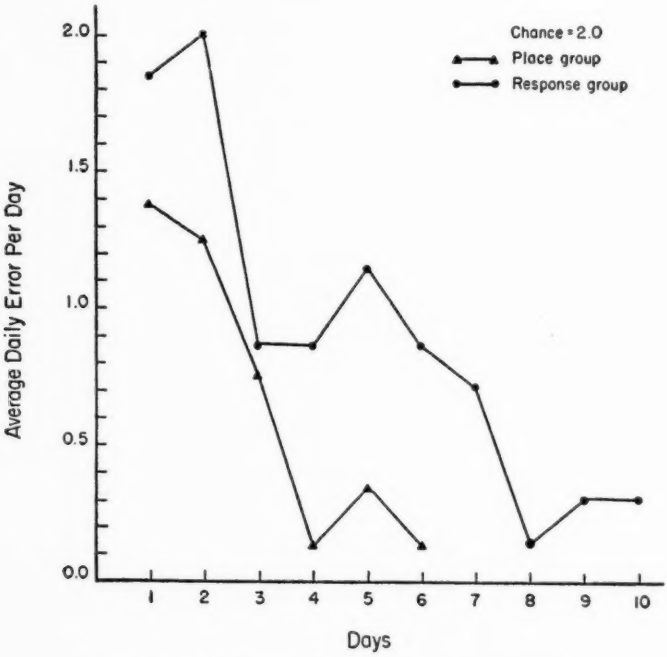

FIG. 2. Average daily errors per day for each group on each day

In the experiment reported in the second article of this series (2), where the correction method was used, it was concluded that "place learning is simpler than response learning." The results obtained for the place-learning group in the present experiment indicate that this conclusion still holds to a considerable degree even when both groups are trained by the non-correction method (cf. Fig. 2). This finding is important. For throughout this series of articles we have maintained that, at least for simple maze problems, a fundamental

factor in learning is that of locating the goal spatially with respect to the extra-maze environment.

The learning of what we have called a 'response-learning group' has at least two possible explanations. (1) The animals of this group could perform correctly if they learned two goal locations; that is, if they learned the spatial location of F_1 with respect to S_1 and that of F_2 with respect to S_2. Or, (2) they could perform correctly if they learned to make the same specific response, e.g., a right turn. At present, our data give us no way of choosing between these alternatives. However, no matter which (if either) is true, it still is significant that such a group will not learn as quickly as a single-place learning group.

Now let us consider the third and fourth questions which this experiment was designed to answer. We asked if animals that had learned to make the same response to get to two different places could maintain that response when the apparatus was rotated 90 degrees around the choice point. We also asked whether animals that had learned to go to the same place from two different starting-places could maintain a consistent response when the maze was rotated through 90 degrees. The data reported do not give an unambiguous answer to these last two questions. A comparison of the two groups on their critical trial, that is, on the trial after the maze had been rotated, gives the following similarities and differences:

1. The animals in neither group showed any marked change in respect to hesitation time or VTE's.
2. The animals in the response group turned correctly, whereas those in the place group did not.[5]
3. The animals in the response group hesitated at a point half way between the choice point and the food cup, whereas those in the place group ran unhesitatingly to the end of the arm into which they had turned.

But this comparison of similarities and differences does not give us easily interpretable answers. However, in any case, we can say that the choice-point cues for the animals of the response-learning group were apparently not the same as those that affected their behavior near and at the goal location. On the other hand, there appears to no such difference for the place-learning group. These findings might enable us finally to discover and state more precisely the differences between what we have called 'place learning' and 'response learning.'

[5] In view of the rather arbitrary determination of what was correct for the place group on the critical trial, it would be more accurate to say that not all of the animals of this group made either a consistent response or went to the same place.

Summary

1. A group of seven rats was run to determine if rats trained by the non-correction method could learn to make the same response to get to two different places. The results were positive.

2. Although the rats of this group learned readily they still did not learn as readily as a place-learning group also run by the non-correction method.

3. After the rats of the response-learning group had learned to make the same response to get to two different places, the maze was rotated 90 degrees to determine whether these animals could maintain their behavior when run with the apparatus in this new position.

4. A similar rotation of the maze through 90 degrees was also carried out with the place-learning group.

5. The behavior of the rats in the two groups after the maze had been rotated through 90 degrees was quite different.

6. The rats of the response-learning group made unhesitating correct turns on the critical trial after the rotation, but stopped about half way out toward the goal and crept forward slowly to the food cup, seized a bit of food and ran towards the other end.

7. The rats of the place-learning group ran without hesitation to the ends of the arms of the T on the critical trial after rotation and showed no disruption. However, only some of the animals ran to the end designated as correct.

8. It is believed that further experiments with such rotation of the maze after learning will eventually permit a further understanding of the difference between what we have been calling 'response learning' and 'place learning.'

(Manuscript received August 1, 1946)

Bibliography

1. Kalish, D. The non-correction method and the delayed response problem of Blodgett and McCutchan. *J. comp. Psychol.*, 1946, **2**, 91–108.
2. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: II. Place learning vs. response learning. *J. exp. Psychol.*, 1946, **3**, 221–229.

THE MASKED THRESHOLD OF PURE TONES AS A FUNCTION OF DURATION *

BY W. R. GARNER AND G. A. MILLER

The Johns Hopkins University and Harvard University

It is strange that in the available literature there is hardly a mention of the relation between the duration of a tone and its threshold intensity. Investigators interested in determining thresholds have not been interested in duration, and investigations of tonal duration have been concerned with reverberation (**6**), loudness (**1**), and differential sensitivity (**3**, **7**), but not with thresholds of audibility. This unseemly gap in our knowledge became of practical importance during the course of certain wartime studies of communication problems, and the authors undertook at that time to fill the gap with the indicated experiments (**4**). These experiments were undertaken to determine the relation between the duration and the threshold for pure tones masked by a noise spectrum.

Above and beyond their practical value, however, the results have a theoretical significance for the time-intensity problem in audition. The purpose of this paper, therefore, is to report these results in a form more available to psychologists, and to develop further the implications of the data for a theory of the relation between duration and intensity.

Apparatus and Procedure

The apparatus for these experiments is shown schematically in Fig. 1. It consisted of two channels, one for sine-wave signals of various durations, and the other for random noise. The outputs of these two channels were mixed, amplified, and led to five sets of headphones for binaural listening. A simple on-off timer turned the tone on for durations ranging from 10 msec. to over 2.0 sec. The timer was self-running, so the signal was repeated at periodic intervals. These short bursts of tone were fed into a set of band-pass filters to eliminate the transients which occur when a tone is suddenly turned on or off. The characteristics of these filters were such that the tone did not reach maximum amplitude until 10 msec. after the tone had been turned on, and likewise did not reach zero amplitude until the tone had been turned off an equivalent time. A 500-ohm attenuator following the filters was calibrated in one-decibel steps, and served as the control attenuator.

The noise circuit consisted of a white-noise [1] generator whose output was filtered by a low-pass filter having a nominal cut-off frequency of 7000 cps. The noise and the tone were mixed by

* This research was done at the Psycho-Acoustic Laboratory under Contract OEMsr-658 between Office of Scientific Research and Development and Harvard University, under sponsorship of the National Defense Research Committee. The authors wish to acknowledge the aid of Prof. S. S. Stevens, Director of the Psycho-Acoustic Laboratory.

[1] 'White noise' is a term used to indicate a noise spectrum in which all frequencies between the stated limits have an equal amount of energy.

means of mixing attenuators (10-db isolation), and the combined signal was then amplified by a high-power, wide-band amplifier. A second attenuator was used to set the overall level of noise and signal. The signals were then led, by means of an LS-33 impedance-matching transformer, to five pairs of dynamic earphones (PDR-8) connected in parallel. An oscilloscope, a VU meter, and a Ballantine voltmeter served as calibrating instruments.

The noise spectrum was determined with a Western Electric Sound Frequency Analyzer, Type RA-277-F. The effective overall band width of the noise was 40 to 5700 cps, as determined by the half-power points of the spectrum. For a uniform band of this width the difference between the overall noise energy and the noise energy in bands one-cycle wide is 37.5 db. The earphones had been calibrated in a six cc. coupler, and had a uniform frequency response over the range of frequencies used. The various signal durations used were determined by measuring the on-time with a chronoscope calibrated in hundredths of a sec. The tones were presented at approximately four-sec. intervals, except that this interval was increased slightly for the longer tones.

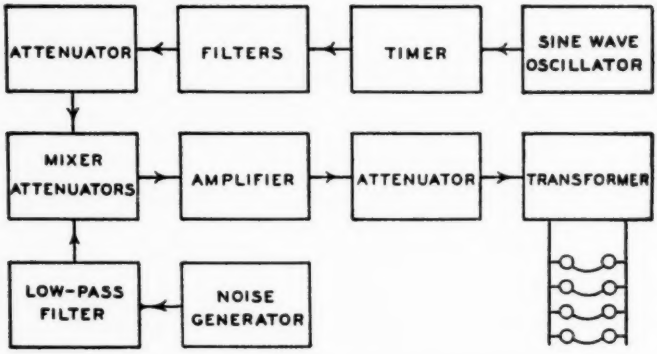

FIG. 1. BLOCK DIAGRAM OF EQUIPMENT

FIG. 1. Block diagram of equipment

In obtaining the data, four *S*s listened simultaneously and binaurally in a sound-deadened room. Before each series of measurements was begun, the levels of both the noise and the signal were calibrated, and the level of the noise was set by means of the step attenuator following the power amplifier. This attenuator controlled the level of both the noise and the signal, so that the ratio of signal strength to noise level could then be read directly from the step attenuator controlling only the tone. When the measures were to be taken, *E*, located with the equipment in a room adjacent to the listener's room, set the tonal signal below the threshold level and increased the signal level one db each time the signal was presented. Each listener had a microswitch which he pressed whenever he heard the tone. This switch controlled a light which *E* could see, and the lights were arranged in such a way that *E* always knew which *S* had pressed the switch. When *S* had heard two signals in succession, the value of the signal-to-noise ratio (S/N) was recorded. *E* continued to increase the level of the tone until all four listeners had heard the tone twice in succession. This procedure was repeated five times, at which time a new series was begun. Since each condition was presented twice, each *S* made a total of 10 judgments for each of the conditions. The measure used as the threshold value was simply the average (in decibels) of all the S/N values recorded for each *S*.

RESULTS

In obtaining S/N measures for minimum detectability, four different frequencies were used: 400, 670, 1000, and 1900 cps. With

each of these frequencies, six overall noise levels were used: 20, 30, 50, 70, 90, and 110 db sound-pressure level. Measurements were also made of the absolute threshold for each of the frequencies, although these measurements turned out to be so unreliable that only the masked threshold measures were used. For each frequency, and for each noise level, eight different signal durations were used: 12.5, 25, 50, 100, 200, 500, 1000, and 2000 msec. Thus, the S/N ratio was obtained for a total of 192 different listening conditions.

There are two ways in which the S/N ratio may be expressed. The first of these is simply to state the ratio of the level of the signal to the overall level of the noise in db. The second and more precise way is to state the ratio of the signal energy to the energy in a one-cycle band width of noise. This latter expression is more meaningful, since an increase in the overall noise energy resulting from an increase in the band width of the noise has little effect on the level at which a tone of a particular frequency will be heard. In this paper, all S/N ratios are expressed in terms of the noise energy in frequency bands one cycle wide.

The S/N ratio at threshold will vary to some extent as a function of the overall noise level. Fig. 2 shows this relation for one frequency

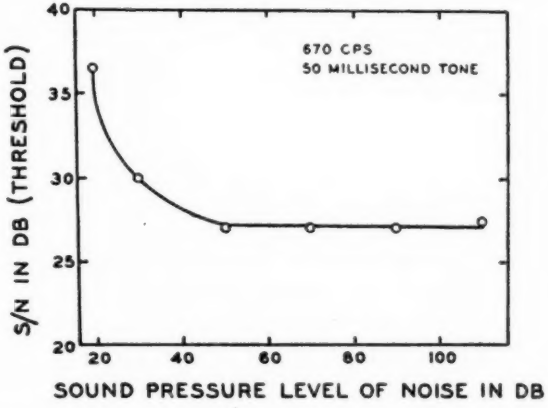

FIG. 2. The relation between the masked threshold of a pure tone and the level of the masking noise.

Noise energy for the S/N ratio is in terms of energy per one-cycle band widths. The reference level for the overall noise is 10^{-16} watts/cm^2. The overall level of the noise is 37.5 db greater than the level per cycle.

and one tonal duration. With very low noise levels, the S/N ratio is larger. As the level of the noise is increased, however, the S/N ratio becomes constant, indicating that the threshold for the signal is

shifted by an amount proportional to the increase in the level of the noise. The curve shown in Fig. 2 is typical of all the curves of that nature obtained in these experiments.

Since the primary purpose of the experiments was to determine the relation between the masked threshold and tonal duration, it was decided to average all the scores for the four highest noise levels. This procedure provided more reliable threshold measures. Table I gives the average threshold values for all the frequencies and durations used. It will be noticed that a value for infinite time is presented in the table. This value is not empirically determined, but has been computed, and the meaning of the computed value will become clearer in the discussion. These same values have been plotted in Fig. 3, with a different symbol for each frequency. Although the points for each of the four frequencies do not fall on the same line, they all present so nearly the same general shape that it seemed permissible to draw one curve through all the points. The curve as drawn has no theoretical significance, but is merely the best visual fit.

TABLE I

THE RELATION BETWEEN TONAL DURATION AND THE SIGNAL-TO-NOISE RATIO NECESSARY FOR MINIMUM DETECTABILITY

Noise is measured in terms of energy per one-cycle band widths.

Duration (msec.)	400 cps	670 cps	1000 cps	1900 cps
12.5	33.9	33.8	31.9	33.1
25	30.7	30.2	29.8	30.8
50	26.9	26.9	26.0	28.7
100	24.1	24.3	23.4	25.8
200	20.2	21.3	20.3	22.7
500	18.5	18.8	18.0	21.3
1000	16.5	17.0	18.4	21.3
2000	16.5	17.1	17.5	20.7
inf.	14.9	15.7	16.8	20.2

The most significant fact about these data is the linear relation between the logarithm of duration and threshold intensity (expressed in decibels) for tonal durations of less than 200 msec. Furthermore, the slope of this line indicates that there is linear integration of acoustic energy up to durations of 200 msec. Thus when the duration is increased by a factor of 10, the S/N ratio is decreased by 10 db, which is also a factor of 10 in terms of acoustic energy or in terms of instantaneous electrical power. The equation for the linear portion of the curve below 200 millisec. is

$$S/N + 10 \log_{10} t = 44.2,$$

where S/N is the signal-to-noise ratio expressed in decibels, and t is duration of the tone in msec.

Above 200 msec., the curve gradually flattens out until there is relatively little change in the S/N ratio when the duration of the tone is increased by fairly large amounts. The integration can be considered effectively complete after about one sec. Any further increase in duration has practically no effect on the stimulus intensity required for minimal audibility.

Discussion

The practical problem is simply the determination of how much loss in sensitivity one ought to expect if the duration of a signal is

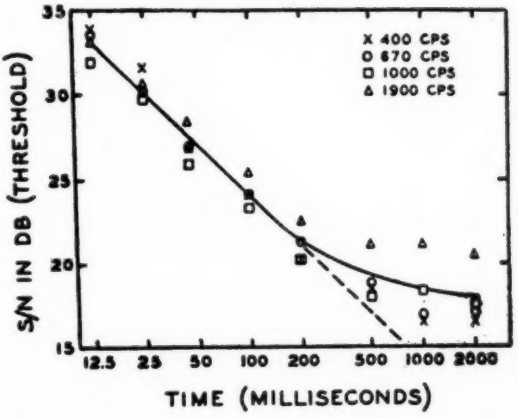

Fig. 3. The relation between the masked threshold of a pure tone and its duration. Noise energy for the S/N is in terms of energy per one-cycle band widths. Each plotted point is the average of measurements for the four highest noise levels, i.e., is the average of 160 observations. The heavy line is a visual fit of the data, and the dashed line is the extension of the curve which represents linear integration.

decreased. The theoretical problem is more than that. What we should like to know is the manner in which the ear integrates acoustic energy. Is it a simple linear integration, occurring only within limited ranges, or is the relation more complicated?

Thus far the data have been presented on log-log plots. An examination of the linear relation between the two variables sheds some light on the general problem. Fig. 4 shows this relation for the 1900 cps tone. Time in this case has been plotted on a linear scale, and the decibel readings for the S/N ratios have been converted into equivalent linear units, the conversion being on the basis of energy, and not pressure. As the time gets shorter and shorter, the stimulus intensity required for the tone to be heard becomes much greater,

and the curve seems to be approaching zero time as its asymptote, with infinite intensity. On the other hand, the threshold becomes lower and lower as the time is increased, but the curve does not seem to be approaching zero intensity as its asymptote. Rather it seems to approach some finite value of stimulus intensity. Altogether, however, the general shape of the curve is certainly that of the usual reciprocal relationship.

From the shape of this curve, it seems quite possible that there is some sort of linear integration, although the integration does not start with any arbitrary zero intensity, but with some value which is effectively zero for the ear. If this is correct, we should no longer expect that the relation between time and intensity would be $It = C$, which is a statement of perfect linear integration.

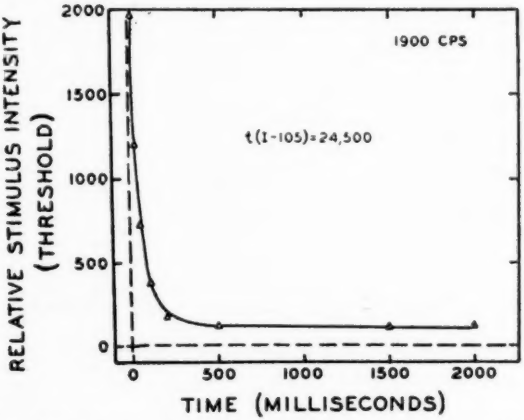

FIG. 4. The relation between the masked threshold of a pure tone and its duration. Both stimulus intensity and time are plotted on linear scales.

Hypothesis: There is certain minimal stimulus intensity (I_0) which is not an effective stimulus for the ear. The ear will, however, integrate all stimulus energy above this minimal value, and this integration is linear with time. Expressed in mathematical form, the following relation should be true.

$$t\,(I + I_0) = C.$$

In this equation, t is linear, and both I and I_0 are expressed in terms of energy.

It should be noted that this equation is of the same mathematical form as that used by Hecht (**5**) to express the same relation for vision, although the general equation is derived on the basis of an entirely different logic. Hecht states the equation in the form $It = C + Dt$,

in which the constant D is small compared to the constant C. This constant D is a result of an adaptation process going on with time, and the quantity Dt will have very little effect when t is small, but will have progressively more effect as t gets larger and larger. Wald (**8**) has also stated a similar equation, on still other grounds, to express the relation between visual area and threshold.

In order to test this hypothesis graphically, several mathematical transformations of the data were required. First, for each S/N ratio, for each of the four frequencies, the equivalent linear quantity (in terms of energy) was computed. With these linear values, the asymptotic value for each of the four frequencies was then computed by the method of least squares. The quantity $I - I_0$ was then determined by subtracting the computed asymptotic value from each linear value of stimulus. If the relation $t\,(I - I_0) = C$ is a true relation, then a log-log plot of the value $I - I_0$ and t should be a straight line, since $\log t + \log (I - I_0) = \log C$. Each value of $I - I_0$ was therefore reconverted into an equivalent decibel scale, and plotted as a function of the logarithm of time. This plot is shown in Fig. 5. Each of the four curves has had its time scale shifted to some extent, so that the intercept constant of the four curves is the same, and they can all be fitted with the same straight line.

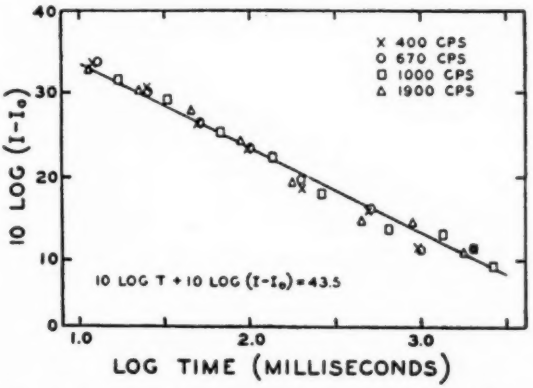

FIG. 5. The relation between the duration of a tone and the masked threshold in terms of $I - I_0$.

This graph is a test of the hypothesis that the ear integrates acoustic energy linearly above a stimulus value which is effectively zero for the ear. The plotted points in this curve are derived from the data of Fig. 3, and each of the four curves for frequency has had its time base shifted to equate the intercept constants.

The straight line drawn in Fig. 5 is the theoretical line which all the points should fit if the hypothesis is perfectly substantiated. Although the fit is far from perfect, it is nevertheless clear that, ac-

cording to this hypothesis, acoustic integration by the ear continues well beyond 200 msec. It should be remembered that in assuming linear integration, there is only one degree of freedom allowed in fitting the data to the curve. The intercept constant is that degree of freedom. If a second degree of freedom had been allowed, in the form of a slope constant, then the fit would have been even more adequate.

Another interesting fact about this particular hypothesis is that the intercept constants for the four curves are quite similar. These constants are (expressed in terms of the logarithmic statement of the equation) 43.6, 43.5, 42.3, and 43.9 for the frequencies 400, 670, 1000 and 1900 cps respectively. In the plotted form, all the data have been corrected so that each curve has the same intercept constant.

Crozier's statistical theory

Crozier's theory (**2**) states in essence that the relation between stimulus threshold intensity and stimulus duration is such that if the reciprocal of I is plotted as a function of the logarithm of t, the resultant curve should be the normal probability integral. This relation is expected on the assumptions that the thresholds of neurons are normally distributed, and that this distribution is reflected when sensitivity $(1/I)$ is plotted as a function of duration. Crozier's theory has seemed adequate for many kinds of visual data, and it seemed possible that these data could be fitted by his hypothesis.

To test that hypothesis the reciprocals of all the data points were computed, and these scores were then changed into percentage scores so that they could be fitted on a probability grid. In order to change the scores into percentages it is necessary to have some value for 100 percent integration, and the reciprocals of the previously computed asymptotic values were used. In order to have all the scores fall on the same curve, it is necessary to have identical slope and intercept constants. With a normal probability integral the intercept constant is equivalent to the mean of the distribution, and the slope constant is equivalent to the standard deviation of the distribution. All time values, therefore, were converted into what are essentailly standard scores so that the mean and intercept constants would be the same.

These corrected data are plotted in Fig. 6, and the straight line drawn there is the theoretical curve which all the data should fit if the hypothesis is substantiated. It can be noticed that this fit is about equally as good as that of the previously tested hypothesis. The main difficulty with this test is that only one or two points go beyond +1 standard deviation in the distribution, and the last six points all fall on the same side of the theoretical curve. Since very little curvature occurs in the normal curve up to one standard de-

viation, these data should not be considered an adequate test of the hypothesis.

It should be pointed out that this test allowed one more degree of freedom in its fit than did the previous hypothesis. Neither the slope of the curve nor the intercept constant were fixed by the hypothesis, so that the data were in effect both rotated and shifted in order to fit the hypothesis. The previous hypothesis, it will be recalled, allowed only one degree of freedom in its fit.

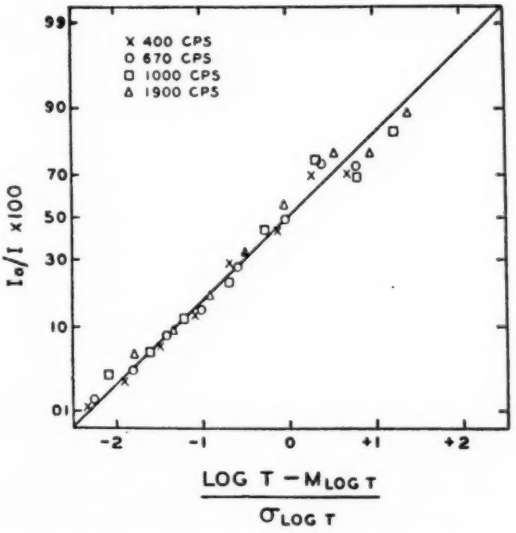

FIG. 6. The relation between the duration of a tone and the masked threshold in terms of probability theory.

This graph is a test of Crozier's statistical theory. The plotted points in this curve are derived from the data of Fig. 3. The reciprocal of I in terms of percentage has been plotted on a probability grid, and the values of time have been changed so that each of the four curves has the same slope and intercept constants.

In graphically testing the hypothesis of a sigmoid fit, it is necessary, of course, to convert sensitivity scores into equivalent percentage scores, and this requires, as mentioned previously, a value for 100 percent. For that value, we chose the reciprocal of the previously computed asymptotes for each of the curves. It might be felt that in using that value, we were not doing full justice to Crozier's hypothesis. We have tried other values for 100 percent sensitivity, however, and the reciprocal of the computed asymptote provides the best fit of these data to Crozier's hypothesis.

Comparison of the two hypotheses

From the data available with these experiments, the two hypotheses seem about equally well substantiated. There are, however, certain mathematical relations between these two curves which should be discussed.

If we have a relation such that $It = C$, and if we plot the reciprocal of I as a function of linear t, then we will have a straight line relation, with the reciprocal of I becoming larger with increases in t. The slope of this line will depend on the constant C. If, however, we plot the reciprocal of I as a function of the logarithm of t, then we will have a monotonic rising function with a positive acceleration, and this curve will resemble greatly the lower part of a sigmoid curve. The top of the curve, under these conditions will continue to rise to infinity.

On the other hand, if we add a constant to I, and then plot the reciprocal of $I + C$ as a function of the logarithm of t, then as I gets smaller and smaller, the value of $1/(I + C)$ approaches the value of $1/C$, and we have provided an asymptote for the reciprocal of I. Then the curve will no longer extend to infinity, but will bend over, approaching the asymptotic value. Thus the curve starts with a positive acceleration, then changes to a negative acceleration with an asymptotic value. This, of course, is exactly what the normal probability integral does.

Thus, any data which fit the equation $t(I - I_0) = C$ will also fit fairly well the requirements of the normal probability integral, since when we plot $1/I$ rather than $1/(I - I_0)$ we have essentially added the constant necessary to provide the curve with an asymptotic value. Data which fit one or the other of these hypotheses will, of necessity, also fit the other hypothesis with some degree of adequacy. On this basis alone, then, it is difficult to choose between the two.

However, the assumption of linear integration for all intensities above an absolute minimum intensity is more stringent in its requirements, and seems simpler in principle than the assumption of an integration with respect to logarithmic time. Until more direct evidence is available, the elaborate and ingenious formulation proposed by Crozier should probably be regarded with a certain cautious conservatism.

Summary

In these experiments, the relation between auditory sensitivity and the duration of a pure tone was measured. Four different frequencies (400, 670, 1000 and 1900 cps), and eight different tonal durations (12.5, 25, 50, 100, 200, 500, 1000, and 2000 msec.) were used. Because of the difficulties involved in measuring the threshold of

hearing in the quiet with any degree of precision and reliability, the masked threshold was determined with a white-noise masking spectrum. The following results and conclusions are presented:

1. The ear integrates acoustic energy linearly up to 200 msec., for when the duration is decreased by a factor of 10, sensitivity is decreased by 10 db.

2. For durations longer than 200 msec., the change in sensitivity is slight, and acoustic integration may be considered complete at approximately one sec.

3. The hypothesis is presented that the auditory system integrates acoustic energy linearly, if the zero intensity is regarded as that inintensity which is effectively zero for the ear. The data substantiate this hypothesis fairly well.

4. Crozier's statistical hypothesis is also tested, and the data fit that hypothesis as adequately as they fit the previous hypothsis.

5. Mathematical relationships between the two hypotheses are shown such that if the data fit one hypothesis, they must of necessity fit the other.

(Manuscript received August 1, 1946)

Bibliography

1. Békésy, G. v. Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. *Physik. Z.*, 1929, **30**, 721–745.
2. Crozier, W. J. The theory of the visual threshold. I. Time and intensity. *Proc. nat. Acad. Sci.*, 1940, **26**, 54–60.
3. Garner, W. R., & Miller, G. A. Differential sensitivity to intensity as a function of the duration of the comparison tone. *J. exp. Psychol.*, 1944, **34**, 450–463.
4. Garner, W. R., & Miller, G. A. The effect of duration on the masked threshold of tones. Part II in: *The masking of signals by noise*, Psycho-Acoustic Laboratory, Harvard University, Oct. 1, 1945, OSRD Report No. 5387. (Available through the Publications Board, U. S. Department of Commerce, Washington, D. C.)
5. Hecht, S. Vision: II. The nature of the photoreceptor process. In: Murchison, C. (Ed.), *Handbook of general experimental psychology*. Worcester, Mass.: Clark Univ. Press, 1934.
6. Lifshitz, S. Apparent duration of sound perception and musical optimum reverberation. *J. acoust. Soc. Amer.*, 1936, **7**, 213–221.
7. Turnbull, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.
8. Wald, G. Area and visual threshold. *J. gen. Physiol.*, 1938, **21**, 269–287.

THE PROBLEM OF STIMULATION DEAFNESS.* II. HISTOLOGICAL CHANGES IN THE COCHLEA AS A FUNCTION OF TONAL FREQUENCY

BY KENDON R. SMITH

Princeton University

The problem of stimulation deafness is one which has maintained a prominent position in auditory research since the latter part of the last century. Earlier work and, to some extent, later work on the problem has been characterized by inexact methods and somwhat equivocal results. Kemp (1) has reviewed the status of the problem as of 1935; since that time, Lurie, Davis, and Hawkins (2) have published further on the nature of stimulation injury.

The first paper in the present series (3) reported a study of cochlear impairment as a function of tonal frequency. In that study, 31 guinea-pig ears were investigated. Each ear was first tested to ascertain its normal cochlear potentials. The ear was then exposed for four min. to a pure tone of either 300 cycles (11 ears), 1000 cycles (8 ears), or 5000 cycles (12 ears) in frequency. The sound pressure for each of these exposure tones was 1000 dynes per sq. cm. After the ear had been exposed, its cochlear potentials were again recorded. It was thus discovered that each ear had suffered a substantially uniform loss of sensitivity throughout the entire frequency-range studied (100 cycles to 10000 cycles), regardless of the frequency of the exposure tone. Moreover, the average loss among the ears subjected to one exposure tone was found to be of about the same order (25 db) as the average loss among the ears subjected to another. The magnitude of loss was found to vary considerably from ear to ear, especially within the 300-cycle group, but little correlation was found between the original sensitivity of a given ear and the degree of loss which it experienced.

It was possible to preserve 14 of these ears for histological analysis: four from the 300-cycle group, four from the 1000-cycle group, and six from the 5000-cycle group. The present paper reports the results of their analysis.

Procedure

Immediately after its electrical tests, each ear was fixed by intravital perfusion through the superior aorta. The fixative was a mixture of Maximow's fluid and 10 percent formalin. After

* This investigation was supported by a grant from the John and Mary R. Markle Fund.

perfusion, the ear was removed from the skull, and fixation was consolidated by immersing the ear first in additional fixative mixture and then in simple 10 percent formalin. The ear was then decalcified and embedded in celloidin.

Each ear was sectioned at 20 micra in the horizontal plane parallel to the modiolar axis. Every fifth section was saved. The retained sections were stained with Ehrlich's hematoxylin and counterstained with eosin. They were then mounted in serial order for analysis.

The ears were inspected microscopically at magnifications up to 430 diameters. To facilitate microscopic analysis, a scale projection map was made of the cochlea of each ear. The map was constructed according to the method of Guild (4). Such a map represents the course of the basilar membrane as it would be projected upon a plane perpendicular to the modiolar axis. The error-in-length produced by such a projection is negligible except for the error involved in projecting the extreme basal end of the basilar membrane. In that region, the basilar membrane hooks sharply, and in a direction roughly parallel to the modiolar axis, about the inner curve of the round window. Accordingly, chords of the hook are measured, and the hook is reprojected with reasonable faithfulness into the plane of the map.

After the cochlear map had been made, the serial sections of the ear were examined successively. At each point at which the cochlear structures were visible, their condition was ascertained. A separate judgment was made upon the condition of each of: (1) Reissner's membrane, (2) the tectorial membrane, (3) each of the hair-cells, (4) the organ of Corti structures other than the hair cells, (5) the basilar membrane, (6) the tympanic lamella (the thin layer of cells adherent to the lower, or tympanic, surface of the basilar membrane),[1] (7) the nerve-supply to the organ of Corti, and (8) the ganglion cells in the ganglion of Corti. The judgments were recorded in standard symbols at appropriate points upon the map. In the procedure followed, independent judgments were made upon adjacent regions of the same structures.

When the analysis had been completed, the condition of the cochlear structures was plotted linearly against distance along the basilar membrane. The linear graph was the final representation of the histological condition of the cochlea.

Results

Considering the intensity of the exposure tones to which the ears were subjected, a surprisingly limited amount of damage was produced. No middle-ear lesions were discovered in any ear. Furthermore, none of the membranes of the inner ear—Reissner's, the tectorial, or the basilar—showed clear-cut signs of damage. The neural tissues of the cochlea all displayed normal cytology; this condition was to be expected, however, because the ears were fixed almost immediately after exposure and there was little opportunity for neural degeneration away from the locus of injury.

When damage was evident, it was confined to the tympanic lamella or to both the tympanic lamella and the organ of Corti. If both structures were injured, they were injured in substantially the same region of the basilar membrane. Injury among the 300-cycle ears was both extensive and profound, and occupied approximately the entire upper half of the cochlea. Among the 1000-cycle ears, lesions were somewhat less extensive and less severe; they centered about two-thirds of the way from the basal end to the apical end of the basilar membrane, leaving structures at both ends of the basilar membrane histologically intact. The damage experienced by the

[1] Lurie, Davis, and Hawkins (2) have called attention to the condition of the tympanic lamella (or 'mesothelial cells') as a possible indicium of cochlear injury.

5000-cycle ears was minor and sharply limited to a region mid-way along the course of the basilar membrane. Within individual groups of ears, and especially for all of the ears considered together, there was little if any correlation between the degree of loss in sensitivity and the general magnitude of lesion.

To illustrate and amplify these general statements, the linear graphs for the injured ears are reproduced in the accompanying figures. Fig. 1 presents the data for the four ears exposed to a tone of 300 cycles. Each ear is represented in that figure by two graphs. In the upper of the two graphs, the ordinate represents the histological condition of the organ of Corti, with the full height of the ordinate denoting complete normality, and lesser heights denoting proportional degrees of normality. In the lower of the two graphs, the ordinate represents the residual proportion of tympanic lamella cells, the full height denoting a 100 percent residual, and lesser heights proportional amounts. In every graph, the abscissa represents distance from the basal end of the basilar membrane, in mm.; the abscissa is truncated at the point at which the basilar membrane ends. The graphs for each ear are accompanied by a parenthetical notation of the average loss in sensitivity suffered by the ear, and the pairs of graphs are presented in order of decreasing loss of sensitivity for their respective ears.

The extensive character of the lesions produced by the 300-cycle tone is obvious in Fig. 1. It is also apparent that the lesions were radical as well as extensive. Those represented in Fig. 1 involved typically the gross displacement from the basilar membrane of the external hair-cells, the Deiters cells, and the external supporting cells; in general, the internal hair-cells escaped obvious injury; the cells of the tympanic lamella were largely stripped from the basilar membrane in the region of damage to the organ of Corti. In every case, the injury occupied the apical 40 or 50 percent of the basilar membrane.

It can be seen that the 300-cycle lesions varied only slightly in extent and severity from ear to ear. In contrast, the respective losses in sensitivity for the various ears ranged approximately from a profound 70 db to a relatively superficial 21 db. Such a lack of correlation between damage and loss, as will be seen, is a general feature of the findings.

Fig. 2 gives similar graphs for three of the four ears exposed to the 1000-cycle tone. The fourth ear was apparently normal histologically. In general, the 1000-cycle lesions centered about a point approximately two-thirds of the way along the basilar membrane, and spared the apical portions, as well as the basal portions, of the basilar membrane structures. Somewhat less extensive than the 300-cycle

lesions, the 1000-cycle lesions were in general also less severe. One ear, No. 443L, displayed no lesions of the organ of Corti. In the two cases in which the organ of Corti was damaged, the external hair-cells

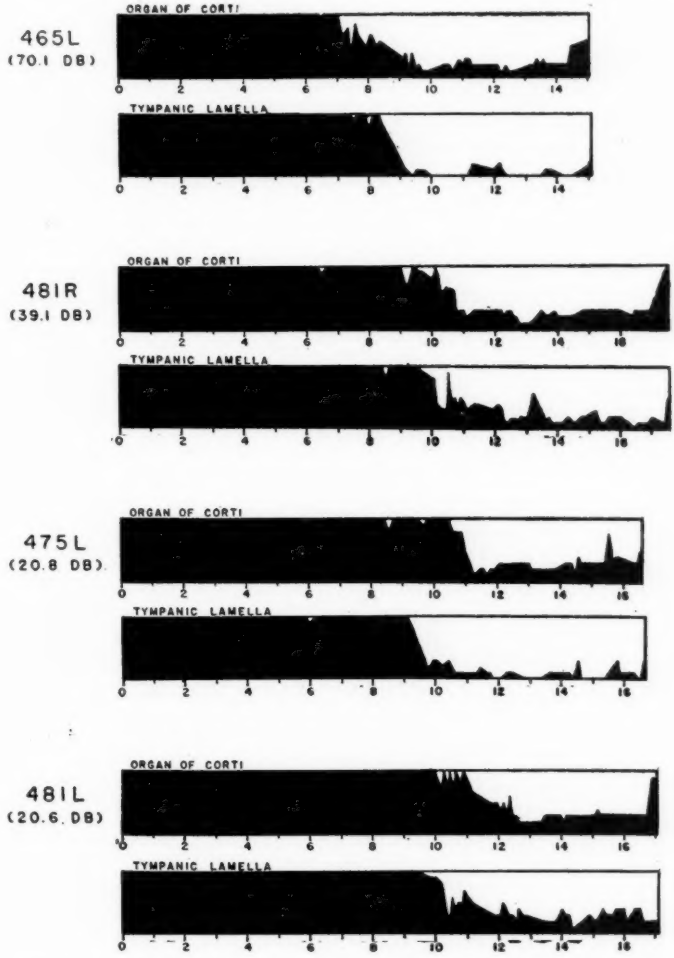

FIG. 1. Lesions produced by 300 cycles. Parenthetical notation gives for each ear the average loss in sensitivity previously found.

were disrupted but not displaced; the remaining structures in the organ of Corti were typically normal histologically.

The lack of correlation between degree of loss and magnitude of lesion is particularly striking in the 1000-cycle group. Ear No. 443L,

which showed damage to only a restricted portion of the tympanic lamella, experienced a loss of approximately 47 db. On the other hand, ears No. 441L and 452L, with obvious and extensive injuries to the organ of Corti and to the tympanic lamella, suffered respectively smaller losses of about 32 and 19 db. The fourth 1000-cycle ear, histologically normal, lost about 10 db in sensitivity.

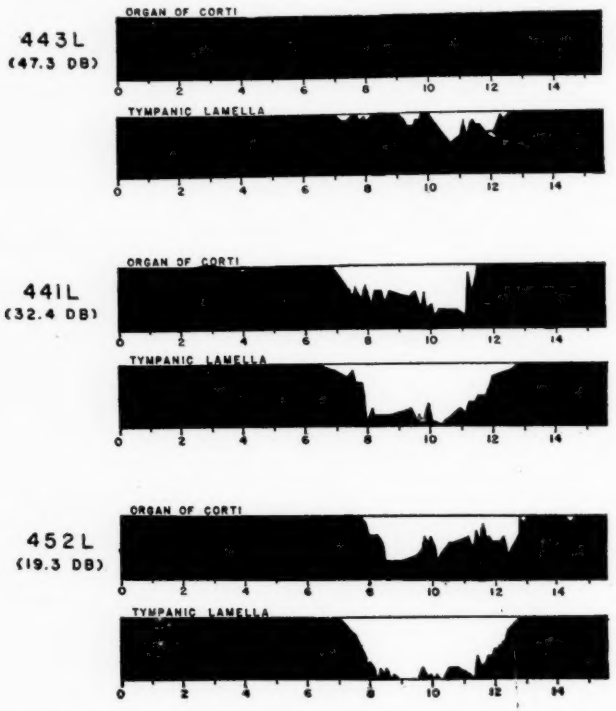

FIG. 2. Lesions produced by 1000 cycles. Parenthetical notation gives for each ear the average loss in sensitivity previously found. A fourth ear (450L; loss, 10.3 db) was histologically normal.

Fig. 3 shows the effects of the 5000-cycle exposure tone upon four ears. The two remaining ears in the 5000-cycle group appeared to be normal histologically. The lesions produced by the 5000-cycle tone were rather sharply localized very nearly at the mid-way point, along the basilar membrane. The injuries to the tympanic lamella and to the organ of Corti were minor, involving in the latter organ merely swellings or minor disruptions in the external hair-cells. Here again, as in the 300-cycle and 1000-cycle groups, there was little

correlation between loss of sensitivity and magnitude of lesion. Ear No. 459R, with at least some damage to the organ of Corti, showed a loss of approximately 15 db; ear No. 458R, with no manifest damage

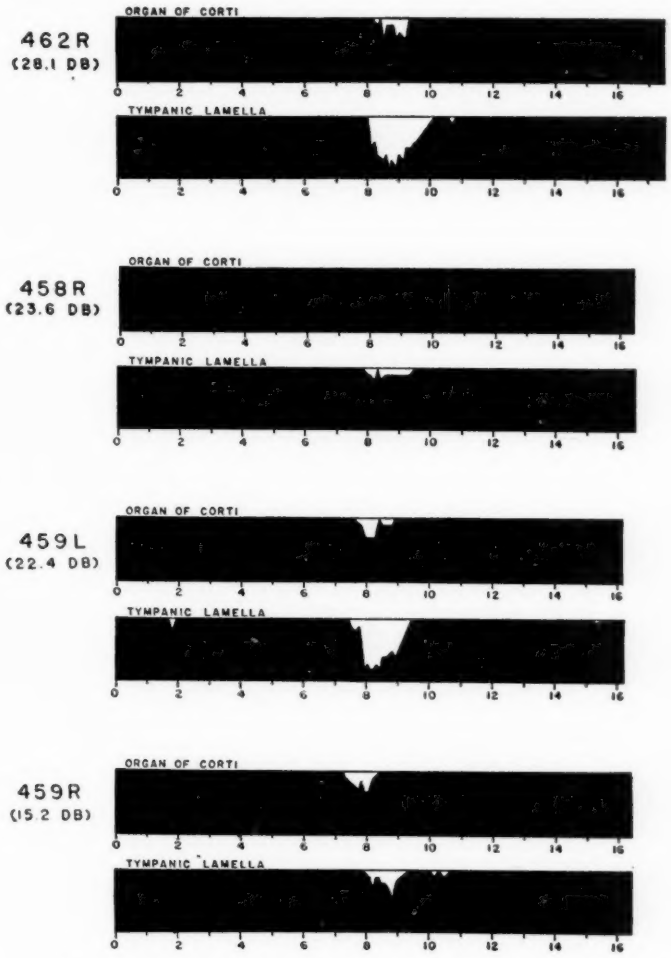

FIG. 3. Lesions produced by 5000 cycles. Parenthetical notation gives for each ear the average loss in sensitivity previously found. A fifth ear (457R; loss, 21.4 db) and a sixth ear (462L; loss, 33.0 db) were histologically normal.

to that organ, lost about 24 db. Indeed, one of the histologically normal ears in the 5000-cycle group (No. 462L) showed the largest average loss in sensitivity (33 db) among all of the ears in the group.

The lack of relationship, from ear to ear, between magnitude of loss and magnitude of damage is most evident when all of the ears are considered together. It appears that the cochlear structures of an ear such as No. 475L or No. 481L (Fig. 1) may be radically and extensively injured, and yet the ear may suffer only a moderate loss (about 21 db); whereas an ear may be apparently normal histolologically (No. 462L, mentioned above) or but slightly damaged (No. 462R, Fig. 3) and still experience a somewhat greater loss (about 30 db). It will be remembered, furthermore, that in the original experimental work upon the ears here considered each of the groups of ears suffered approximately the same average loss in sensitivity. In contrast, it now appears that the average extent and severity of lesion varied greatly and systematically from group to group.

The systematic variation in magnitude of lesion with frequency of exposure tone is in itself a matter of interest. The general nature of this variation is shown not only in the figures already presented but also in the three plates. Plate 1 is a photomicrograph of a typical 300-cycle lesion, taken in the area of injury in ear No. 481R. The severity of damage is obvious. Plate 2 is a photograph taken in the area of injury in ear No. 441L, and it shows the relatively moderate damage inflicted by the 1000-cycle tone. Finally, the minor damage wrought by the 5000-cycle tone is shown in Plate 3, a photograph of a portion of the damaged region in ear No. 459L.

Discussion

The number of cases involved in the present study is not great. The systematic nature of the lesions, however, is suggestive of certain conclusions.

The first is that tones of different frequencies primarily affect different areas of the basilar membrane. The focus of action for the 5000-cycle tone in the present experiment was evidently a point virtually half-way along the course of the basilar membrane; 5000-cycle injury was sharply localized at that point. The 1000-cycle tone injured a wider area than did the 5000-cycle stimulus; however, it is probable that the area of primary response to the tone lay approximately in the center of the injured region, and thus about two-thirds of the way along the basilar membrane. The 300-cycle tone created still more widespread damage; furthermore, the 300-cycle lesions 'ran off' the apical end of the basilar membrane to a certain extent. The locus of maximal response is thus most difficult to judge in the case of the 300-cycle tone, but it probaby lay in a region about eight-tenths of the way from base to helicotrema. It is noteworthy that these localizations agree very closely with those obtained by Lurie, Davis, and Hawkins (**2**) with similar techniques.

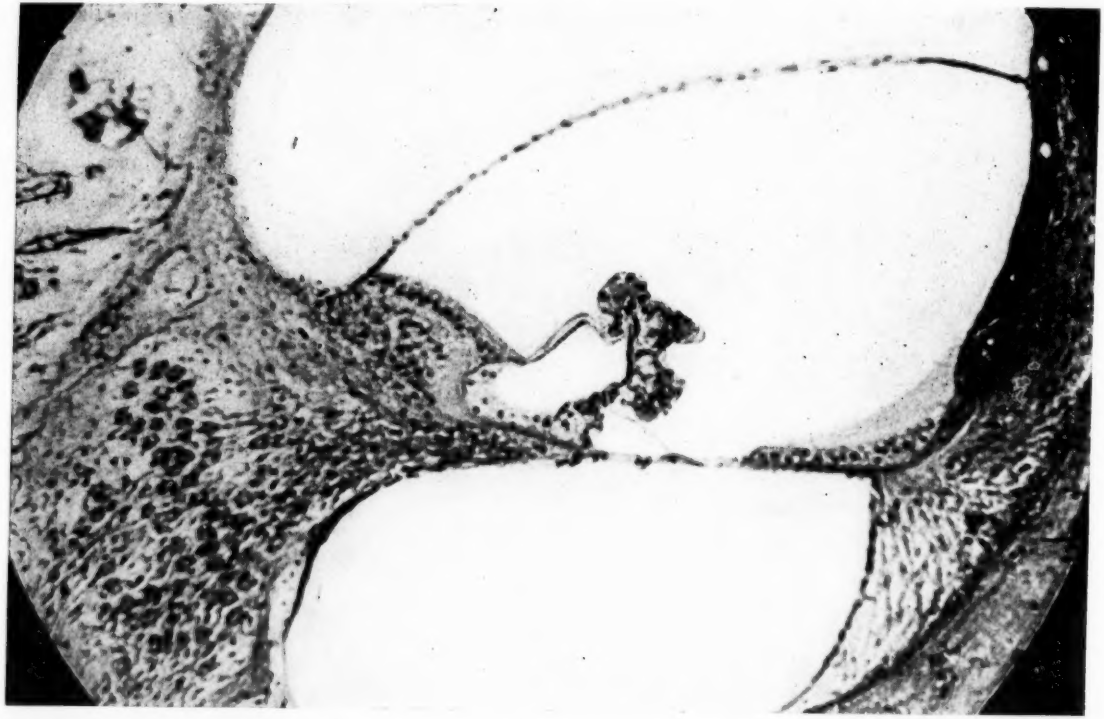

PLATE 1. Typical 300-cycle injury. 228×. Gross displacement of external hair-cells, Deiters cells, and external supporting cells; tympanic lamella largely missing. Ear No. 481R.

PLATE 2. Typical 1000-cycle injury. 228×. Disruption of external hair-cells; tympanic lamella largely missing. Ear No. 441L.

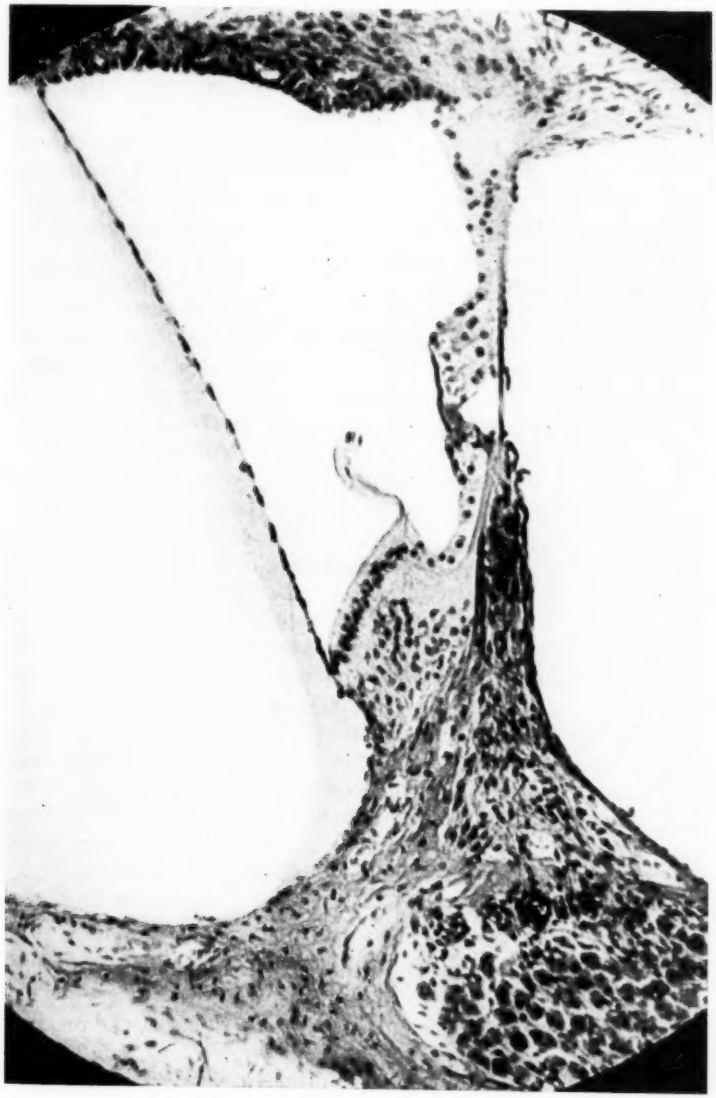

PLATE 3. Typical 5000-cycle injury. 228×. Minor disruption of external hair-cells; tympanic lamella largely missing. Ear No. 459L.

Whether these foci of action, found for extremely intense tones, are also the foci of action for less intense tones of the same frequencies is another question. It is conceivable that the locus of maximal response is to some extent a function of intensity, and that results obtained from exposure experiments are thus not directly comparable with results obtained by other techniques. For whatever reason, however, it is apparent that the loci of action found in the present study are somewhat at variance with those found by other techniques (**5**, **6**, **7**). Notably, the 5000-cycle area in the present case is located much farther than it is otherwise reported from the basal end of the basilar membrane. Usually placed within the basal half-turn of the guinea-pig cochlea (**5**, **6**, **7**), the 5000-cycle area was found here approximately at the junction of the first and second turns. On the other hand, both the 1000-cycle area and the 300-cycle area were found displaced slightly downward in the cochlea. The net effect is to revise the usual frequency-map of the cochlea to restrict the frequencies below 5000 cycles to a considerably shorter apical range of the basilar membrane than is usually postulated, with the region devoted to the frequencies from 5000 to 300 cycles being especially limited and the region devoted to frequencies below 300 cycles being somewhat expanded.

A second conclusion is also suggested: if sound pressure is held constant, stimuli of lower frequency produce observable lesions of greater magnitude than do stimuli of higher frequency. This relationship may indicate that the structures responsive to low frequencies are more fragile than those responsive to high frequencies. Such differential fragility is by no means certainly demonstrated, however. The complex relationship among the several dimensions of the stimulus enforces certain systematic variations in the stimulus, and it is possible to account for the observed variability in lesion in terms of variation in stimulus, and without an assumption of differential fragility.

In a given medium, in this case the air, a sound stimulus may vary basically in four dimensions:

(1) Frequency.
(2) Particle-displacement (amplitude), which is independent of frequency.
(3) Particle-velocity, which is directly proportional to the product of amplitude and frequency.
(4) Particle-acceleration, which is directly proportional to the product of amplitude and frequency-squared.

The relationship among these variables is such that, although individually they may be held constant, if any one is varied, two others

must also vary concomitantly. In an attempt to determine the etiology of cochlear destruction, one might choose frequency as his constant and the three positively-correlated aspects of intensity as his variables; he would then no doubt come to the somewhat truistic conclusion that increased destruction follows increased intensity. To perform a finer analysis, therefore, one must at least vary frequency. Intensity may be held constant in one of its dimensions, but it will vary perforce in at least two of those dimensions.

In the present experiment, frequency of exposure tone was varied. Intensity was held constant in the medial dimension of particle-velocity by controlling a linear variant of particle-velocity, sound-pressure. Since particle-velocity is proportional to the product of amplitude and frequency, this product was a constant; and it follows from this constancy that amplitude was *inversely* proportional to frequency. Finally, particle-acceleration varied *directly* with frequency, because particle-acceleration was proportional to the constant product (amplitude times frequency) times frequency again.

The nature of the variation in each of the dimensions of the aerial stimulus is thus known. It seems probable that these variations were followed reasonably well by corresponding variations in cochlear action; that is, as frequency of exposure tone was increased, not only did the frequency of cochlear vibration follow the frequency of the stimulus, but, simultaneously, amplitude of vibration was decreased, and acceleration was increased.

Of these relationships, the inverse association between the independent variables, frequency and amplitude, is of particular interest. It has been seen that a similar inverse relationship existed between frequency and magnitude of lesion. It appears then that as frequency of exposure tone was increased, amplitude was decreased, and severity and extent of lesion decreased correspondingly. The manifest suggestion is that the gradation in lesions was due to the gradation in amplitude—that the effect of changing amplitude was great enough to override the effect of a contrary change in frequency.

A final problem is posed by the results of the present study: in general, changes in the histological condition of the cochlear structures are not reflected systematically by changes in the electrical audiogram. Although a given area of the basilar membrane seems to be particularly concerned with a certain frequency, there is no sign that observable injury to a restricted area is associated with a restricted loss in responsiveness; rather, sensitivity to a broad range of frequencies is reduced quite uniformly. Furthermore, a small lesion in a limited portion of the basilar membrane may or may not be associated with a smaller general loss than will a large lesion in the same area. It would appear, then, that the observed lesion is not the

entire injury to the cochlear apparatus, but that a general submicroscopic injury is engendered by over-stimulation.

Summary

A group of 14 guinea-pig ears, previously exposed to intense pure tones, was examined histologically to ascertain the nature of the damage produced. All observable injury was confined to the tympanic lamella or to both the tympanic lamella and the organ of Corti. Lesions produced by a 5000-cycle tone were minor and sharply localized half-way along the course of the basilar membrane. Lesions produced by a 1000-cycle tone were considerably more severe and extensive, and centered about two-thirds of the way, from base to apex, along the basilar membrane. Injury due to a 300-cycle tone was very severe; the injury involved virtually the entire upper half of the cochlea and appeared to center approximately eight-tenths of the way from base to apex. There was no evident correlation, from ear to ear, between magnitude of lesion and magnitude of loss in electrical sensitivity (previously determined). Furthermore, previous determinations had shown general, rather than restricted, electrical losses for all the ears.

The results suggest a revision of the current frequency-maps of the cochlea to restrict the frequencies below 5000 cycles to a considerably shorter apical range of the basilar membrane than is usually postulated, with the region devoted to the frequencies from 5000 cycles to 300 cycles being especially limited and the region devoted to frequencies below 300 cycles being somewhat expanded. It is pointed out that the graduation in magnitude of lesion as an inverse function of frequency can be explained by the hypothesis that amplitude of cochlear vibration is primarily responsible for observable injury, as well as by the hypothesis of differential fragility of the cochlear structures. The lack of correlation between the histological effects and the electrical effects of over-exposure suggests that over-exposure produces, in addition to the observable lesions, widespread invisible injury to the cochlear structures.

(Manuscript received August 27, 1946)

References

1. Kemp, E. H. A critical review of experiments on the problem of stimulation deafness, *Psychol. Bull.*, 1935, **32**, 325–342.
2. Lurie, M. H., Davis, H., & Hawkins, J. E., Jr. Acoustic trauma of the organ of Corti in the guinea pig. *Laryngoscope*, 1944, **54**, 375–386.
3. Wever, E. G., & Smith, K. R. The problem of stimulation deafness. I. Cochlear impairment as a function of tonal frequency. *J. exp. Psychol.*, 1944, **34**, 239–245.

4. Guild, S. R. A graphic reconstruction method for the study of the organ of Corti. *Anat. Rec.*, 1921, **22**, 141–157.
5. Culler, E. An experimental study of tonal localization in the cochlea of the guinea pig. *Ann. Otol., etc.*, 1935, **44**, 807.
6. Stevens, S. S., & Davis, H. *Hearing.* New York: John Wiley and Sons, 1938.
7. Culler, E., Coakely, J. D., Lowy, K., & Gross, N. A revised frequency-map of the guinea-pig cochlea. *Amer. J. Psychol.*, 1943, **56**, 475–500.

AGE AND SEX DIFFERENCES IN CRITICAL FLICKER FREQUENCY *

BY HENRYK MISIAK

Fordham University

A. Introduction

The repeated stimulation of the retina by flashes of light at a certain rate produces a phenomenon called flicker. At a very slow rate each flash can be perceived separately. With an increasing rate a coarse flicker is experienced at first, then a fine flicker, and when the rate is further increased the flashes become superimposed and the intermittent light is perceived as a steady illumination, a stage that is called fusion. The rate at which this fusion is just produced is termed the fusion point. Since the flashes of light come at a regular rate, separated by a period of no-light or darkness, they are spoken of as cycles, each cycle having two phases, a light phase and a dark phase. The number of cycles per sec. necessary to produce fusion is called fusion frequency limen or critical flicker frequency (c.f.f.). The latter term will be used hereafter.

There are few visual phenomena that have been more extensively studied than flicker and fusion. Interest in this field started about 180 years ago, although with passing years this took on many different aspects. The chief object of recent flicker studies has been the determination of c.f.f. and the factors upon which c.f.f. depends. These factors influencing c.f.f. can be divided into physiological and psychological, peripheral or central, or a combination of all of them. The physiological or peripheral factors constitute the best known field. Less known are the psychological or central, sometimes called dynamic factors.

Recently age has also been stressed as an important factor in the study of c.f.f. There have been studies that covered the age range from 5 to 80 years, indicating no increase in c.f.f. from childhood to adulthood but a definite decline in the value of c.f.f. after 40 years of age. Koffka, making an inference from the apparent analogy between fusion in vision and tetany produced by intermittent stimulaation, suspected lower c.f.f. in infants than in adults (**16**, pp. 62-63).

* This paper is part of a dissertation presented to the Graduate School of Fordham University in partial fulfillment of the requirements of a Ph. D. degree. The author wishes to express his thanks and appreciation to Dr. B. R. Philip and Dr. J. F. Kubis under whose mentorship the research has been carried out.

In 1931 Meili and Tobler (**18**) in an experiment with eight children between five and six years of age, and five adults of unspecified age, failed to find a significant difference in c.f.f. between adults and children. In 1936 Riddell (**23**), using five *S*s, found a decrease in c.f.f. with age.

In 1934 Hartmann (**11**) used 30 young children and 30 adults with normal visual acuity and normal color discrimination. The ages of the children ranged from six to 11, of the adults from 18 to 25. The apparatus consisted of a disc of two 180° cardboards. The disc was rotated by a D.C. motor with regulated speed which could be read off at any given instant from an engineering tachometer coupled to the shaft of the motor. At first the author found higher frequency for children, but in another series of experiments no differences were observed, and this led him to the conclusion that the basic mechanisms involved in c.f.f. mature relatively early (**11**, p. 126). He found, however, higher c.f.f. in males of both young and adult groups. From the computed average deviations less variability was indicated in adults than in children, and less in females than in males.

In 1941 Simonson, Enzer and Blankstein (**25**) used 47 *S*s varying in age from 10 to 80. They divided the *S*s into four groups: group one (10–19 years of age) contained four *S*s, group two (20–29) 18, group three (30–39) 10, and group four (over 40) 15. Although this last group had, according to the authors, *S*s as old as 80 years of age, no measures of central tendency or variability of age are given. The sexes were represented in equal numbers. The apparatus consisted of a rotating disc that interrupted a beam of light from an electric bulb. The speed of the rotation was measured by a mechanical revolution counter. The authors claim high accuracy of the obtained measures, indicating that the recheck in the same experiment showed a variation that exceeded only in exceptional cases 11.5 flashes, whereas the recheck on different days showed the variations did not exceed ±2 flashes unless the *S*'s condition was changed. They obtained an average value of c.f.f. for all groups of 44.9 cycles per sec. In the analysis of their results they considered separately minimum, maximum, and average values of each group and found a decrease in maximum and average values but not in minimum values, and a decrease in variability with age.

In 1942 Miller (**22**) attempted to throw some light on the question of c.f.f. as a function of chronological age or maturity, of sex differences, diurnal variations and practice effect, and the possible effect of loss of sleep on c.f.f. In his experiment flicker was produced by a beam of light interrupted by a disc connected to a reducing friction-disc drive which in turn was driven by a synchronous motor. Two scales permitted readings to hundredths of a revolution and were found consistent within an error of ±0.5 percent. The speed of the disc was varied by means of the synchronous motor and was independent of the variations of the electric current. The *S*s composed two groups, 44 boys, ages from 7–11 to 18–7, and 34 girls from 6–0 to 18–11. In the experiment the *S*s were set for the disappearance of flicker which they signaled by means of a telegraph key, and this point was taken as the c.f.f. value. The *E* started with the speed much below the fusion threshold and gradually increased the speed until the *S* signaled the disappearance of flicker. At each sitting five readings were taken and each *S* was tested on at least three different days.

The results of this experiment can be summarized as follows: the range of c.f.f. extended from 31.6 to 73.0 cycles per sec.; there was some evidence, though apparently not consistent or reliable, for an increase of c.f.f. with age. The boys showed higher c.f.f. than the girls in each age group, a fact that would corroborate Hartmann's findings. The variability of the readings decreased with sittings but the average c.f.f. increased from the first to the second sitting. This is interpreted by the author as practice effect.

In 1945 Brozek and Keys (**3**), in search for a good test of the normal degenerative processes in man that could be used in gerontological studies, found the flicker test a promising and useful tool. Their interest was in finding answers to two questions: whether the average c.f.f. level changes with increasing age, and whether there is a significant association between age and the change in c.f.f. at the end of the workday. The apparatus had a flashlight bulb, the light of which was interrupted by a rotating disc driven by a motor.

Fifty-six women ranging in age from 18 to 60, divided into four age groups, 18–25, 26–35, 36–45, 46–60, with 19, 17, 8 and 12 individuals respectively, served as *S*s. The readings were

taken in the first and the last hour of the working day with 42 *Ss* at clerical tasks and 14 *Ss* at microscopic work. Both upper and lower c.f.f. thresholds were determined. The c.f.f. value was the average of four consecutive determinations made in each session. The mean values of c.f.f. for the above mentioned groups were 46.70, 45.74, 45.39, 40.92 with standard deviations 4.18, 3.59, 2.89, 3.27. A significant difference at the five percent level between the mean of the 45–60 group and the means of the other groups was obtained, but none among the other groups. As to the 'general fatigue' as allegedly manifested in lower c.f.f. at the end of the day's work, no relationship between age and fatigue was found. The authors stress the value of using c.f.f. measures in gerontological studies and the necessity of further investigations in this field.

It has been felt that a more extensive study with reference to the difference in the c.f.f. between youth and old age ought to be carried out. In view of the relatively small number of *Ss* used and of the fact that actually 60 years of age was the limit in previous investigations, it seemed worth while to test larger age groups and to have the old age groups range from the age of 60 to as far upward as possible. Several questions suggested themselves to the author of the present study. If there is a decrease in mean values of c.f.f. with age, will it hold true for both eyes as well as for single eye? What is the variability of the performance of young and old? Will there be sex differences?

The two primary hypotheses of research considered in this paper are:

1. There is a significant difference in the c.f.f. value between young and old people;
2. There are sex differences within the age groups.[1]

B. Method and Procedure

I. Apparatus

The main concern of an experimenter determining the c.f.f. of an *S* is to assure that this c.f.f. is a true measure and is a result, as far as the technique goes, of the factors he controls. If all the factors, like intensity of light, size of the test-patch, frequency of cycles per sec., light-dark ratio and the like, are well controlled, he can have confidence that the obtained c.f.f. is characteristic of the visual mechanism of the *S* and a result of physiological and psychological agents inherent in this particular *S*. Hence, the necessity of effort to improve the technique, especially in the apparatus, so as to control all variables that are known to influence c.f.f. Many devices were used in order to secure the best control of these variables of which the most important seem to be: 1. the intensity of light, 2. the frequency of cycles per sec. and the ease with which it can be changed and read, 3. the right foveal or peripheral stimulation of the retina. The effectiveness of the control of these variables is a criterion of the value of the apparatus and the experimental set-up.

With a few minor changes the apparatus used in the present study was built from the specifications and data presented in an article by Henry (12). It was an electronic apparatus which had a three-watt neon glow lamp giving an intermittent light at a rate determined by three adjustable resistors. Another neon lamp, similar to the lamp producing flicker, served as a voltage regulator, thus assuring the constancy of the flicker rate, even if the supply voltage varied. A small cathode ray oscilloscope tube was built in the apparatus.

The intensity of light was constant and could not be changed. The flash duration of the cycles could be changed from $\frac{1}{3}$ to $\frac{2}{3}$ by adjusting a tap on a resistor. The frequency of cycles

[1] The results refering to the monocular and the binocular c.f.f. and to the variability of the performance of the individuals will be given in subsequent papers.

was adjusted by three resistors, two serving as a coarse adjustment, one as a fine adjustment. There were 23 steps in the coarse adjustment resistors (C.R.), each step representing a change of an average of three cycles, the total range of the resistor being from about 15 to 60 cycles per sec. (c.p.s.). Fine adjustment resistor (F.R.) was a continuously variable resistor, its range being on the average three cycles. Just as is any electronic apparatus, this instrument was affected by the change of the current from supply line, warming up, external temperature and the like. These factors, if not controlled, would upset the rate of flashes and make an accurate assessemnt of c.f.f. extremely difficult. A potentiometer of 500 ohms served as a calibrator to eliminate these difficulties. After the position of the dials for each frequency was established, it was possible to bring the apparatus to the desired norm by slight adjustment of this calibrator. That was done by setting the dials for such frequencies as 30, 45, 60 and the others, as they were known and established from a previous calibration, and by observing Lissajous figures appearing on the screen of the cathode ray oscilloscope. The Lissajous figures for those frequencies are simple and easily recognizable. Correcting the calibration for one frequency meant a correction for all frequencies.

The basic calibration of the apparatus was guided by observing Lissajous figures on the screen of the cathode ray oscilloscope of the apparatus and comparing them with the Lissajous figures on a separate standard oscillograph which was connected with the flicker apparatus and a twin oscillator at the same time. The latter Lissajous figures were the product of combined oscillations, namely from the flicker apparatus and from the twin oscillator. By this means the following frequencies were experimentally found and subsequently used for the calibration curve: 24, 25.7, 31.6, 30, 33.3, 35, 36, 37.5, 40, 42, 42.8, 45, 46.7, 48, 50, 51.4, 52.5, 53.3, 55.6, 60. From the curve all the readings in terms of the dials, as found during the determination of c.f.f., could be easily translated into corresponding frequencies. It is obvious that the determination of frequencies based essentially on the observation of the cathode ray oscilloscope, is more exact than those obtained by the devices determining the speed of a rotating disc. By the use of this method, exactness better than a tenth of a cycle could be obtained.

The following are some technical data that are necessary for the proper evaluation of the c.f.f. of the *Ss*.

The test-patch was formed by a circular aperture of a black tube put in front of the flickering neon lamp. The *S* looked through the eyepiece into the tube with the circular aperture at its end.

The diameter of the test-patch was 5 mm., the area 19.6 sq.mm. The distance from the filament of the neon lamp to the test-patch was one in. The total distance from the neon lamp to the eye of the subject was 13.8 in. The subtending visual angle was 48 min., assuring foveal observation of flicker when the eyes were fixated upon the test-patch.

At the surface of the neon lamp the intensity of light was 15 foot-candles; one in. from the lamp filament, at the circular test-patch, the intensity was six foot-candles. A lightmeter with a photonic cell (Weston, model 614) was used for these measures.

The proportion of light phase to dark phase in a cycle was 0.6. This was found by measuring the light and dark phases of the figure obtained on the screen of an oscillograph when the latter was connected with the flicker apparatus, using the sweep current.

The apparatus was equipped with an eyepiece with two holes through which the *S* observed the test-patch. This eyepiece was made in such a way that no other light could come into the *S*'s eye except that of the test-patch.

II. Subjects

One hundred *Ss* were tested, and their scores constitute the data for computation and discussion. Quite a large number of *Ss* were rejected before the test or during the test because they manifested some visual abnormalities like amblyopia, color blindness, diplopia or were too nervous and unable to respond to the instructions.

The 100 *Ss* formed two groups equal in number: young and old. The young group of 25 men and 25 women ranged in age from 19 to 30 years; the old group, also of 25 men and 25 women ranged from 63 to 87. Table I shows the range, mean, median and SD's of the age. It is evident that the means and variabilities for the two sexes were well equated within each age group.

The young group was composed of 41 college and graduate students of Fordham University and 9 Carmelite nuns. The old group was composed of 32 inmates of St. Patrick's Home for the Aged, Bronx, New York City, and 18 inmates of the Home for the Aged of the Little Sisters of the Poor, 106th Street, New York City.

Before the experiment the following tests were administered:

TABLE I

DISTRIBUTION OF AGE OF THE FOUR EXPERIMENTAL GROUPS

Age groups	Young		Old	
Sex groups	Male	Female	Male	Female
Number of *Ss*	25	25	25	25
Range	18.9–29.8	19.7–29.8	63–86	66–87
Mean	22.8	23.2	73.8	73.7
Median	21.7	22.3	73.0	73.0
SD	3.30	3.15	6.30	5.01

1. *Ocular dominance.*—A Parson's manuscope was used. This simple test, although not perfect, shows a high reliability (**19**). 69 percent were right-eyed and 31 percent left-eyed, which is in agreement with the findings of Miles (**19**) and Crider (**5**).

2. *Color blindness.*—The Ishihara test was used (**13**). All dichromates and anomalous trichromates were rejected.

3. *Visual acuity.*—This was tested with Snellen chart from the distance of 20 feet. Both eyes were tested, then the left and right eyes separately. If the *S* wore glasses, the readings were taken with and without glasses. A great variety of the degree of visual acuity, from normal vision 20/20 or better to 20/70 or worse, was found. No *S* was rejected on account of bad sight unless this was due to cataract, injury to the eye, previous operation or any eye disease. In the course of the experiment no evidence was found of relationship between visual acuity and c.f.f. An investigation of the cases where one eye was weaker than the other and there existed at the same time a difference in c.f.f. between the two eyes showed that in 20 instances the stronger eye showed a higher c.f.f., and in 19 instances the weaker eye showed a higher c.f.f. There were a number of instances where a marked divergence in visual acuity between the right and left eyes was not accompanied by the divergence in c.f.f. values.

4. *Blood pressure.*—Old people had their blood pressure taken by a trained nurse with a standard sphygmomanometer. This was done because it was felt that it might have some bearing on c.f.f. (**24**). However, neither diastolic nor systolic blood pressure seems to have any effect on the individual differences in c.f.f. The correlation coefficient between systolic blood pressure and c.f.f. for both eyes for the old men was .10, for the dominant eye .22, neither value being significant at the five percent level.

All *Ss* were fairly healthy. The old people, as might be expected, were subject to various ailments characteristic of old age, but there was nothing in their medical records that would indicate any disease which could affect the c.f.f. If there was any suspicion of this nature, those *Ss* were eliminated from the experiment. The young as well as the old *Ss* were very cooperative and apparently well motivated to do their best in the experimental situation.

III. *Experimental procedure*

Each *S* was tested on three consecutive days, each day having one session. The first session was regarded as a practice session and was devoted to a careful explanation of the mode of perceiving the phenomenon, so as to insure that flicker and fusion would be actually observed and properly discriminated. Although the general procedure was the same in all three sessions, only the scores from the last two sessions were used in computation. The first session will be referred to hereafter as a practice session, the others as the first and second sessions or days.

The *S* was seated in front of the apparatus. The seat was so adjusted in height and position with regard to the apparatus that the *S* could easily look through the eyepiece of the apparatus at the test-patch. The eyepiece had a shield similar to those with which automatic movies in amusement arcades are equipped. This shield prevented outside light from coming into the *S*'s eye and as his forehead leaned against it, the distance from the test-patch to the eyes was kept constant. The eyepiece had two holes, one for each eye. When monocular c.f.f. readings were taken, one hole was screened with a piece of black cardboard. In this way both eyes were tested in succession, for example, the left eye, then both eyes, then the right eye. The placing of a screen in the hole for the eye that was to be eliminated made the observation of flicker possible

in an easy and speedy way. Moreover, this seemed to have and advantage over devices applied directly to the eye, as it avoided irritation of the eye and delay. In the described arrangement the *S* very often did not realize whether he was using both eyes or only one eye, or whether it was the right or the left one.

The room was darkened so that no light from without could penetrate. At a distance of about 4–5 feet from the subject there was a three-watt neon lamp, the same kind of lamp which was producing flicker in the apparatus. The *S* was well light-adapted when he entered the room. After two min. of adaptation to the light of the neon lamp, the experiment started. In the practice session an obvious flicker was presented and the *S* was asked: "Do you see the bright spot? Do you see that the light flickers?" When it was certain that the *S* saw flicker, the frequency was increased so that the light looked steady (beyond 60 c.p.s.): "Now it does not flicker any more, does it?" The *S* was instructed not to move his head or blink his eyes when looking at the test-patch. For the actual determination of c.f.f. the following procedure was used. First, an obvious flicker was presented and when seen by the *S*, C.R. was switched one step higher increasing the frequency about three cycles. If flicker was still noted by the *S*, the resistor was again switched step by step until the *S* did not see flicker any more. This critical step was recorded and the resistor increased by two steps so as to assure absence of flicker. From that point the resistor was switched down step by step until the *S* saw flicker again. This step was also recorded. The process was repeated two or three times, sometimes more, if the answers were inconsistent. In the next two experimental sessions it was never repeated and the readings were taken as reported by the *S*. The *S* could not see the dial and could not know whether the dial of the resistors was switched up or down.

The above procedure secured the determination of limits, one of obvious flicker and one of apparent fusion between which lay the actual c.f.f. On the dial of C.R. these limits always represented one step and this corresponded to a difference of about three cycles. For example, flicker may have been seen at 19 (frequency 44.42) and no flicker at 20 (frequency 47.24). Having these limits, the lower step of C.R. was set (in the above example it would be 19) and the flicker presented to the *S*. And now F.R. was used which during the shifting of C.R. was kept at zero. These instructions were given to the *S:* "Do you see flicker? Look at the bright spot and say 'Now,' when you do not see flicker any more." As soon as he said so, F.R. was turned to the extreme right end, increasing the flicker frequency beyond the apparent c.f.f. of the *S*, and then slowly and gradually turned back, until the point at which the *S*, according to further instructions ("Say 'Now' when you see flicker again"), reported flicker. The frequency was immediately lowered to a strong flicker by turning the dial of R.F. to the extreme left and then slowly increased until the disappearance of flicker, then the frequency increased again and slowly decreased until the reappearance of flicker. The critical points of disappearance and reappearance of flicker were recorded. After the four readings were taken, a one min. rest was given, and then the other eye was tested according to the prearranged rotating scheme which is presented below. The whole procedure in which those four readings were taken lasted about two min. After three sets of four readings, that is, after both eyes together, the right and left eyes separately were tested, a rest of two min. was given during which the calibration of the apparatus was checked. The whole session took from 20 to 30 min.

Each session, then, consisted of six sets of four readings, two of them indicating the points at which flicker disappeared and two of them, the points at which flicker reappeared. Each set was started with coarse change of frequency by means of C.R. until no flicker was reported. Then the resistor was turned back until flicker appeared. Between these two limits F.R. was used to determine the exact points at which flicker appeared or disappeared. As is evident, the *S*s were set, not for fusion, but for flicker throughout the whole experiment, which is important to remember in the light of the experiments of Riddell (**23**) and Knox (**14, 15**). The latter states that c.f.f. "is slightly higher when the subject is set for flicker" (**15**, p. 129).

The order of testing the eyes was as follows:

Practice Session and Session or Day I	1. Left eye.	2. Both eyes. 2 min. rest	3. Right eye.
	4. Right eye.	5. Both eyes.	6. Left eye.
Session or Day II	1. Both eyes.	2. Right eye. 2 min. rest	3. Left eye.
	4. Left eye.	5. Right eye.	6. Both eyes.

The time of the sessions for the old *S*s was the morning between 8:30–12 o'clock. For the young *S*s, morning and early afternoon. In most cases the *S*s were tested at the same time of the

day on three consecutive days. For example, if an *S* was tested on the first day at 10 A.M., he was tested the next two days at the same time.

As to the *Ss*' errors of refraction, a uniform practice was kept throughout in that, if an *S* wore glasses constantly, he observed the flicker with the glasses on. Artificial pupils were not used. The difficulty in using them with the old people, the small test-patch, its short distance from the eye and the specific form of the eyepiece justified dispensing with the artificial pupils.

C. Results

The basic data can be presented in an overall table, Table II, based upon 4800 individual readings made on 100 *Ss* on two consecutive days. It has to be remembered that in each of four groups there are 25 *Ss*. Each of the means, such as 48.01, is based upon the c.f.f. scores of 25 *Ss*, the score of each *S* being the average of 16 readings, eight obtained on the first day, eight on the second. Four of these eight readings were the determination of the upper threshold (from flicker to fusion) and four of the lower threshold (from fusion to flicker). The SD of the scores of 25 *Ss* (still using testing condition B) is 4.62. The range of the c.f.f. in this group is 39.24–53.78.

TABLE II

Means, Standard Deviations and Ranges of the Average C.F.F. for the Four Groups

Age groups	Young						Old					
Sex groups	Male			Female			Male			Female		
Test cond.	B	D	N	B	D	N	B	D	N	B	D	N
Mean	48.01	43.77	43.84	46.18	43.48	43.45	37.04	34.93	34.38	39.58	37.11	35.33
SD	4.62	3.40	3.62	5.15	5.06	5.17	5.14	6.56	5.90	4.99	4.53	5.10
Range from	39.24	36.67	36.85	38.81	37.29	36.93	28.27	18.29	27.54	30.05	29.64	28.60
to	53.78	52.47	51.88	56.25	55.64	55.91	53.49	52.79	53.64	54.76	43.49	43.56

Testing conditions: B—both eyes; D—dominant eye; N—non-dominant eye.

I. *Sex differences*

A rough inspectional analysis shows that there is no large sex difference between the c.f.f.'s for any of the B, D, or N categories within the young and old groups. However, the *t*-test is the proper way of evaluating the difference.

TABLE III

The *t*-Values for Sex Differences within Age Groups

	Age groups compared	
Testing condition	Young (men vs. women)	Old (men vs. women)
Both eyes	1.317	1.776
Dominant eye	.238	1.597
Non-dominant eye	.269	.609

$t_{.05} = 2.010$ for 48 df.

Table III presents the *t*-values obtained in testing the significance of the differences between the means of both sexes for young and old groups. The differences for the B, D, and N categories were tested. Thus, 1.776 is the *t*-value obtained from testing the difference between the binocular c.f.f. means (B) of the old men and old women. This value of *t* is not significant since $t_{.05}$ for 48 degrees of freedom is 2.010. No other *t*-value in the table is significant when the five percent level of significance is used as the criterion for the rejection of the null hypothesis.

II. *Age differences*

Table IV gives the *t*-values obtained from the test of significance of the differences between the means of young and old groups. The men of the young group were compared with the men of the old group, and the women of the young group with the women of the old group. Every *t*-value is significant at the one percent level, since the $t_{.01}$ for 48 d.f. is 2.683.

TABLE IV

THE *t*-VALUES FOR AGE DIFFERENCES WITHIN SEX GROUPS

	Sex groups compared	
Testing condition	Male (young vs. old)	Female (young vs. old)
Both eyes	6.415	4.615
Dominant eye	5.973	4.684
Non-dominant eye	6.855	5.600

$t_{.05}$ = 2.010, $t_{.01}$ = 2.683 for 48 df.

D. Discussion

The present investigation differs from all other studies that dealt with the determination of c.f.f. in the use of a new apparatus producing flicker.[2] Thus far, a motor driven disc was used in all the investigations. In this investigation a totally new device has been applied, where the source of flicker is a lamp itself electrically adjusted and the number of cycles per sec. is estimated by an electronic arrangement. As already mentioned the apparatus was originated by Henry (**12**), and he made the preliminary experiments with it on 16 *Ss*. Since the purpose of Henry's experiments was the comparison of the results obtained by the use of the electronic apparatus with those obtained by other apparatus, a sort of check, and reported in only a few lines, the present investigation can be regarded as a first venture into the field of flicker study utilizing an electronic apparatus.

[2] In an article by Graybiel *et al.* (**10**) there is mention that an 'electronic oscilloscope' was used, but there is no description of it. The article is short and the information about the experiment is very scant.

I. *The range and means of the c.f.f. values*

In the determination of c.f.f. either two or one eye can be used. Some investigators determined c.f.f. when both eyes were used in the observation of flicker, like Creed and Ruch (**4**), Simonson *et al.* (**25**), Miller (**22**), some when only one eye was used like Knox (**14**), Henry (**12**) and Ross (**24**). There are other authors who do not give any information as to whether one or two eyes were used (**3**, **10**), or, if one was used, they do not say which one. The present investigator determined c.f.f. for each eye separately and for both together, keeping all data for separate analyses in order to determine whether age or sex differences would exist under one or more (or even all) of the testing conditions herein designated as B,D,N. Another aim was to throw some light on the relationship between monocular and binocular c.f.f. The latter problem will be discussed in another article.

It is interesting to compare, first of all, the range of c.f.f.'s and mean c.f.f. obtained for two age groups of this study with other studies, although it has to be borne in mind that the conditions of various studies were more or less, different, and therefore it is difficult to draw any definite conclusions. The difficulty of comparison has been pointed out by Ross (**24**).

In view of Bartley's findings concerning the light-dark ratio (**1**, pp. 120-121), the c.f.f.'s obtained in this study, because the light intensity was medium and the light-dark ratio employed was approximately equal, should be expected to reach their maximum. The range for the young group was 36.67–56.25, for the old group 18.29–54.76 c.p.s. (Table II). The range of c.f.f.'s of the old group cannot be compared with the range obtained by other investigators, since if any of them tested old people, the number of *S*s was too small and the limit of age lower than in this study. For the young group this comparison is easier, as the groups compared are nearly equal in age. Simonson, Enzer and Blankstein (**25**) report a range 40.8–54.1, and Brozek and Keys (**2**) 36.50–56.00. A striking similarity between the latter range (with 56 *S*s) and the range found in this study is to be noted.

There can be some speculation as to the highest c.f.f. ever obtained for the human eye. Hecht, Shlaer and Verrijp place it between 45 and 53 (**1**, pp. 122-123); in the study of Gellhorn and Hailman (**9**) the highest was 51; in the research of Simonson, Enzer and Blankstein (**25**), it was 54.1. Brozek and Keys (**2**) reached the highest, 55.75 and 56, which is almost identical with the 56.25 found by the present investigator. Miller (**22**) reports c.f.f.'s as high as 59.7, 61.9, and even 73.0, but these values can be subjected to criticism, mainly because of the unsatisfactory calibration of scale and pre-

ponderance of the upper threshold in the determination of c.f.f. due to the specific experimental procedure (**22**, p. 22).

The survey of mean c.f.f.'s reveals still more resemblance between the findings of this study and those of others. The means c.f.f. is generally in the vicinity of 45 (**2**, **3**, **7**, **12**, **22**, **25**, **26**). The mean for the young group of this study is 47.10 for both eyes, 43.63 for the dominant eye and 43.65 for the non-dominant eye, which is in agreement with 46.70 for the 18-25 age group and 45.74 for the 26-35 age group of Brozek and Keys (**3**), with 45 of Henry (**12**), with 45.60 of Miller (as calculated from his data, **22**, pp. 28-29), with 45.9 of the 10-19 age group and 46.2 for the 20-29 age group of Simonson, Enzer and Blankstein (**25**).

II. *Sex differences*

Speaking *a priori* one can hardly see why there should be sex differences in a phenomenon like flicker perception. It does not differ essentially from any other sensorial perception, and, if no sex differences exist in the perception of auditory, cutaneous, or kinaesthetic stimuli, one would be inclined to infer that no differences should be found in flicker perception and consequently in c.f.f. values. Yet these differences were reported by Miller (**22**) in children in favor of boys and by Hartmann (**11**) in children and adults also in favor of males. Starting the experiment, the present investigator was convinced of the existence of sex differences in c.f.f. and hence the positive form of his hypothesis regarding sex differences set up before the experiment.

The facts reflected in the statistical analysis of the c.f.f. values of both age groups force one to question seriously the hypothesis of sex differences. No sex differences were manifested within the threefold criterion exemplified by the three testing conditions, both eyes, dominant and non-dominant eye. Superficial examination of the means of the c.f.f. values of the groups shows that the three means of the young female group, i.e., for both eyes, dominant and non-dominant eye, are smaller than the corresponding means of the young male group, whereas in the old age groups the opposite is true (Table II). When the c.f.f. values of the young men were compared with those of young women, and old men with old women, and the *t*-test applied all of the differences proved to be not significant (Table III).

It can be seen that these results are at variance with the results of Hartmann (**11**) and Miller (**22**), who reported sex differences. No apparent explanation is available. However, some shortcomings of both experiments make one question the results. In Hartmann's experiment the apparatus and the control of speed of the rotation

were inadequate. There was also a lack of control of light and dark adaptation, fixation of the eyes and size of the pupil. As to Miller's experiment, the criticism is directed to the rather crude inter- and extrapolations of the calibration, and to the procedure which unduly overemphasized the flicker-fusion threshold.

III. *Age differences*

Hartmann (**11**) compared 30 children of six to 11 years of age with 30 adults of 18 to 25 years and found no significant differences in c.f.f. Miller (**22**) tested 78 *Ss* of age ranging from six to 19 years and could not prove any significant change in c.f.f. with age. Simonson, Enzer and Blankstein (**25**) used 47 *Ss* of 10 to 80 years of age and found a significant drop in c.f.f. after 29. Lee (**17**), whose primary study was fatigue as indicated by a change in c.f.f., did not find marked difference with age for 275 *Ss* whose age ranged from less than 30 to over 40 (only 34 *Ss* were over 40). Brozek and Keys (**3**) studied 56 *Ss* ranging in age from 18 to 60 and reported a distinct decrease in c.f.f. in the late forties. Led by the reasons explained before, the present investigator set up two groups considerably apart in age in order to test the hypothesis of decrease in c.f.f. with age. It may be noticed that while most investigations were limited to about 60 years of age, this investigation resumed work at this age and went beyond 80 years of age.

The results presented in Tables II and IV prove that there is a considerable decrease in average c.f.f. in old age. Young people compared with old people do not manifest significant differences in variability among individuals or in variability in individual performance. They do manifest, however, highly significant difference in average c.f.f. values. This is evident in all three testing conditions. The differences exist also in the range of c.f.f. values or in minimal and maximal values between the young and old groups, as can be seen from Table II. From this table it is evident that the range of the old group is almost twice that of the young group and while the minimum values differ considerably, the maximum values differ comparatively little. This is not what Simonson, Enzer and Blankstein (**25**) found; in their study, values did not differ much for all four age groups except for the maximum values. Brozek and Keys (**3**) found slightly different results. The decrease in maximum values was less pronounced, but there was also a decrease in minimum values. Had these investigators tested more *Ss*, it is possible that the results would have been different. Neverthless, in the present study the maximum values of the old age group are somewhat puzzling. It just happened that there was only one man and one woman who obtained such high c.f.f. values, the former consistently in all three

testing conditions, the latter only for both eyes. The difference between each of these values and the next highest in the old group was about eight cycles. The present investigator, being suspicious of the possibility of some error, tested these two *Ss* again and again and always found that their c.f.f. was above 50. No clue for the explanation could be found in the personality of these *Ss*. The lowest c.f.f. also looks exceptional among the other c.f.f.'s, for the next lowest is about six cycles higher. Yet, a low c.f.f. in old age is more probable than a high one. However, when there is no real justification for regarding these cases as exceptions, they have to be accepted as they were found in the experiment, and hence the fact remains that the range of c.f.f.'s in old age in the group tested in this experiment is almost twice as large as in youth. It must, however, be remembered that comparisons made with ranges might lead to erroneous conclusions due to the relative unreliability of this measure.

In connection with the cases just discussed it will be well to remember that according to Miles 10 to 25 percent of men in later maturity and old age "are able to do as well in most tests as the average in the middle maturity or more efficient group" (**20**, p. 357). In visual acuteness a group of men in a 50–69 age-group had 21.6 percent equal or better scores than the average mean for groups 18–29 and 30–49 years of age combined (**21**). A similar situation exists in the present results. Many old people had as high c.f.f.'s as the young people, even *Ss* around 80 years of age having higher c.f.f.'s than some of the *Ss* in the young group.

In conclusion it can be regarded as proven that c.f.f. decreases with age but to establish the form of the curve of this decrease, further research is indispensable, especially in the years between 40 and 60. Simonson, Enzer and Blankstein (**25**) had only 15 *Ss* in this range, Lee (**17**) had 34, and Brozek and Keys (**3**) about 20. Investigations with larger numbers of *Ss* will be able to establish when the decrease in c.f.f. begins. On the basis of the published results (**3**, **25**) such a decrease is noticeable after 30 but significantly only after 40 years of age.

What causes the decrease in c.f.f. in old age? Many important changes occur in the visual mechanism in old age (**8**). The changes are manifested mainly in the retina, in the optic nerve, and in the visual functions themselves. In the retina senile macular degeneration sets in, colloid bodies or drusen develop on Bruch's membrane, and sclerotic changes are found in the retinal blood vessels. In the optic nerve local arteriosclerotic changes cause senile atrophy of this nerve. Presbyopia as the main variation of the visual function, starting around 44 years of age, is manifested in the loss of accommodation due to sclerosis of the crystalline lens and partly to a gradual

weakening of the ciliary muscles (**6, 28**). There is no high correlation of other visual functions with c.f.f. (**3**), and thus the decrease in c.f.f. points rather to the atrophy of the optic nerve or even to the atrophy of the cerebrum. Whatever may have been the errors of refraction of the old group tested in this experiment, they could not affect much the observation of flicker, partly because the *Ss* wore glasses and partly because the small visual angle probably excluded refraction completely.

Recently a great interest has been shown in gerontological studies (**27**). Brozek and Keys look upon the c.f.f. test as a 'useful tool' in these studies (**3**). They see it as satisfying such criteria as the ease of administration, freedom of the effects of practice and consistency in successive testing sessions (**3**). The present investigator agrees with this view, but would like to draw attention to the following facts evident in his research:

1. In old age there is an appreciable variability among individuals with respect to c.f.f. values, but this is not significantly different from the variability of the young group.[3]
2. High c.f.f. values, sometimes higher than average performance of young age, can be found in old age.
3. There is a considerable overlap of the c.f.f. distributions for the young and old *Ss*.

The flicker test can serve as a useful tool in gerontological studies if these facts are kept in mind. A c.f.f. value can never be an index of senile degeneration unless the measures have been taken over several years and the decrease of c.f.f. value has been observed.

E. Summary and Conclusions

One hundred *Ss*, 50 from 19 to 30 years of age and 50 from 63 to 87 years of age, with the sexes equally represented in the two groups, had their foveal critical flicker frequency determined for both eyes together, dominant eye, and non-dominant eye. The c.f.f. values were determined on three consecutive days. Each c.f.f. was based on eight readings from flicker to fusion and eight from fusion to flicker.

The flicker was produced by a three-watt neon glow lamp of an electronic apparatus. The light from this lamp entered the eye of the *S* through a circular aperture five mm. in diameter subtending a visual angle of 48′. The rate of flicker was controlled by resistors with accuracy better than a tenth of a cycle per sec. The brightness of the test-patch was six ft. candles. The apparatus was equipped

[3] This point will be discussed in detail in another paper.

with a small cathode ray ocilloscope. By means of the Lissajous figures appearing on the screen of this oscilloscope the dials of the resistors were calibrated and the rate of flicker could be controlled.

All the *Ss* were free from any pathological eye condition. The ocular dominance was determined with Parson's manuscope. Before the test the *Ss* were adapted to a light coming from the three-watt neon glow lamp. Under experimental conditions they viewed the test-patch from a distance of 12.8 in.

The conclusions which have been reached are as follows:

1. There is a significant difference in the c.f.f. values between young and old age for both eyes, dominant and non-dominant eye.

2. There are no sex differences in c.f.f. values, in either the young or the old groups.

3. The decrease in c.f.f. in old age is probably due to the degeneration of the optic nerve and cerebrum.

4. The range of the c.f.f. values is greater in old age.

From a general consideration of all factors it seems that the c.f.f. test can be useful in gerontological studies.

(Manuscript received July 29, 1946)

BIBLIOGRAPHY

1. BARTLEY, S. H. *Vision.* New York: Van Nostrand, 1941.
2. BROZEK, J., & KEYS, A. Flicker-fusion frequency as test of fatigue. *J. industr. Hyg.*, 1944, **26**, 169–174.
3. BROZEK, J., & KEYS, A. Changes in flicker-fusion frequency with age. *J. consult. Psychol.*, 1945, **9**, 87–90.
4. CREED, R. S., & RUCH, T. C. Regional variations in sensitivity to flicker. *J. Physiol.*, 1932, **74**, 407–423.
5. CRIDER, B. A battery of tests for the dominant eye. *J. gen. Psychol.*, 1944, **31**, 179–190.
6. DUKE-ELDER, W. *Textbook of ophthalmology*, St. Louis: C. V. Mosby Co., 1936. Vol. I.
7. ENZER, N., SIMONSON, E., & BLANKSTEIN, S. S. The state of sensory centers in patients with hypothyroidism. *Ann. internal Medicine*, 1941, **15**, 649–665.
8. FRIEDENWALD, J. S. The eye. In: E. V. COWDRY (Ed.), *Problems of ageing.* (2nd Ed.) Baltimore: Williams and Wilkins, 1942, Chapter 18.
9. GELLHORN, E., & HAILMAN, H. The effect of anoxia on sense organs. *Proc. Amer. Soc. exp. Biol.*, 1943, **2**, 122–126.
10. GRAYBIEL, A., LILIENTHAL, J. L., JR., & HOWITZ, O. Flicker fusion tests as a measure of fatigue in aviators. *J. Aviat. Med.*, 1943, **14**, 356–359.
11. HARTMANN, G. W. Comparison of the flicker thresholds in children and adults. *Child Developm.*, 1934, **5**, 122–126.
12. HENRY, F. An electronic apparatus for testing fatigue by the visual flicker method. *J. exp. Psychol.*, 1942, **31**, 538–543.
13. ISHIHARA, S. *Tests for colour-blindness.* (9th Ed.) Tokyo: Kanehara, 1940.
14. KNOX, G. W. The relation of dynamic factors in flicker and fusion. Unpublished. Doct. Diss., Ohio State Univ., 1940.
15. KNOX, G. W. Investigations of flicker and fusion. *J. gen. Psychol.*, 1945, **33**, 121–154.
16. KOFFKA, K. *Growth of the mind.* New York: Harcourt-Brace, 1924.
17. LEE, R. H. Critical fusion frequency of flicker. In: JONES, R. R., *et al. Fatigue and hours of service of interstate truck drivers.* U. S. Public Hlth. Serv. Bulletin, No. 265, 1941, Washington, D. C.

18. Meili, R., & Tobler, E. Les mouvements stroboscopiques chez les enfants. *Arch. Psychol.*, Genève, 1931, **23**, 131–156.
19. Miles, W. R. Ocular dominance demonstrated by unconscious sighting. *J. exp. Psychol.*, 1929, **12**, 113–126.
20. Miles, W. R. Abilities of older men. *Person. J.*, 1933, **11**, 352–357.
21. Miles, W. R. Psychological aspects of ageing. In: E. V. Cowdry (Ed.), *Problems of ageing*. (2nd Ed.) Baltimore: Williams and Wilkins, 1942, Chapter 28.
22. Miller, V. L. The critical frequency limen for visual flicker in children between the ages of six and eighteen. *Genet. Psychol. Monogr.*, 1942, **26**, 3–55.
23. Riddell, L. A. Use of flicker phenomenon in investigation of field of vision. *Brit. J. ophthal.*, 1936, **20**, 385–410.
24. Ross, R. T. The fusion frequency in different areas of the visual field. I. The foveal fusion frequency. *J. gen. Psychol.*, 1936, **15**, 133–147.
25. Simonson, E., Enzer, N., & Blankstein, S. S. The influence of age on the fusion frequency of flicker. *J. exp. Psychol.*, 1941, **29**, 252–255.
26. Simonson, E., Kearns, W. M., & Enzer, N. Effect of oral administration of methyltesterone on fatigue in eunuchoids and castrates. *Endocrinology*, 1941, **28**, 506–512.
27. Stieglitz, E. J. *Report of a survey of active studies in gerontology*. Unit on Gerontology, Division of Chemotheraphy, U. S. Publ. Hlth. Serv., 1942.
28. Stieglitz, E. J. *Geriatic medicine*. Philadelphia: Saunders, 1943.

THE INFLUENCE OF SIZE ON PREFERENCES FOR RECTANGULAR PROPORTION IN CHILDREN AND ADULTS

BY WALTER C. SHIPLEY, PRISCILLA E. DATTMAN, AND BARBARA A. STEELE

Wheaton College, Norton, Massachusetts

Orientation and Background

The object of this experiment was to proceed a step further, along the line followed by Thompson (1), in investigating age differences in preference for rectangular proportions. While his investigation was limited to the study of constant length (but varying size), ours provided also for analysis of the influence of the size factor.

He had four chronological-age groups—preschool, third grade, sixth grade, and college—rank rectangles on the basis of aesthetic preference. The rectangles, 12 in all, were of black cardboard and uniform length of $2\frac{3}{8}$ in., ranging in width-length ratio from .25 to .75, with a duplicate at .40, by ratio steps of .05. Among other things his data revealed certain characteristic differences between the preferences of the two school-age groups—the third and sixth grade—and the college group. (The preschool group showed no consistent preference for any of the rectangles.) In terms of median preference ranks the school-age groups—except for a slight inversion between the two narrowest rectangles—showed an increasing preference for the wider rectangles. The college group, on the other hand, and this is in line with the general findings of other investigators working with adults showed an increasing preference for rectangles only up to width-length ratio of .55, though with continued high preference for ratios .60 and .65 (in the region of the 'golden section'), followed by a marked dropping off in preference for the higher ratios.

It occurred to us that inasmuch as the size, i.e., area, of the rectangles increased with the width, the *S*s might have been influenced in their judgments by size, as well as by proportion, especially as they had been instructed simply to pick the rectangle which they liked best, next best, etc. The present experiment, which employed two series of rectangles—one with length, and the other with area, held constant—was designed to investigate this possibility.

Procedure

Materials, presentation and directions.—The materials consisted of 12 black cardboard rectangles, a white card to display them on, and a covered box to deposit them in. The rect-

angles fell into two series of six each, with width-length ratios of .25, .35, .45, .55, .65, and .75, respectively, corresponding to alternate ones of Thompson's series. The Series A rectangles were of uniform length—2⅜ in., as were Thompson's; the Series B rectangles were of uniform area—2.82 square in., corresponding with the mean area of his series and also our Series A. In presenting the rectangles, all 12 were arranged in random manner on the white card. Our procedure, except where already pointed out, was identical with that of Thompson (1, p. 51) which we cite verbatim:

> Each *S* was given the following instructions: "Look at *all* of these (*E* gestured toward rectangles and paused for three seconds). *Which one do you like best?*" As soon as a choice was made *E* placed the rectangle *S* liked best in a covered box and continued,"*Now which one do you like best?*" This procedure was followed until each *S* had made the [12][1] possible choices.
>
> The older *S*s frequently asked questions about the purpose of the experiment when the cards were presented to them. *E* always evaded answering these questions and repeated the instructions, "Which one do you like best?" None of the *S*s was told the purpose of the research until all of the data had been collected since such knowledge might have become general enough to have affected the responses of new *S*s.

Subjects.—Two groups of *S*s served in the experiment—a group of 100 women college students and a group of 100 school children.[2] The college students ranged in age from 17.6 years to 24.7 years, with a mean of 19.7 years, and σ of 1.13 years. The school children ranged from 8.3 to 16.2 years, with a mean age of 11.0 years, and σ of 1.77 years. The composition of the school-children group is shown by grade, sex, and age, in Table I.

TABLE I

Composition of School-children Group by Grade, Sex, and Age

Grade	*N*		Range of CAs	Mean of CAs	σ of CAs
	Male	Female			
3rd	18	8	8.3–12.1	9.4	.97
4th	13	11	8.9–15.1	10.6	1.39
5th	8	10	9.3–13.7	11.1	1.05
6th	5	16	10.8–15.2	12.1	1.06
7th	3	8	12.0–16.2	13.5	1.35

Results

Median preference ranks for each width-length ratio, for each series and each group, are presented graphically in Fig. 1. The solid-line curves represent the A series (constant length), and the broken-line curves, the B series (constant area). To facilitate in distinguishing significant trends from chance variability in comparing the curves, Table II has been drawn up. It lists, for each pertinent pair of curves, the points along the abscissa at which the divergence is statistically significant at the $<.01$ level. Comparison of the curves reveals the following trends:

[1] 'Eleven' in Thompson's procedure.

[2] We wish to thank Mr. Laurence G. Nourse, Superintendent of Schools, and Miss Loretta McHugh, Principal of the Norton Center School, Norton, Mass., for their cooperation in making the school children available to us.

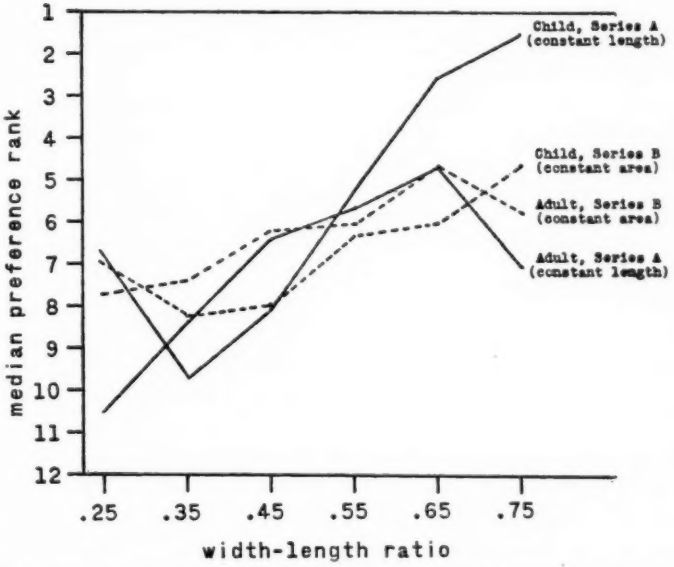

FIG. 1. Curves showing median preference ranks for children and adults (college students) for rectangles of different width-length ratios. The solid-line curves show the preferences for the constant-length rectangles, and the broken-line curves, for the constant-area rectangles. A rank of '1' indicates 'best liked.'

1. The two A curves (constant length) are very similar to the corresponding curves presented by Thompson. Except for the initial dip (and a minor similar dip was suggested in his data), the A curve for the school children shows a continuously increasing preference for the wider rectangles, while the A curve for the adults (college students) shows an increasing preference only up to width-length ratio .65, and then a striking falling off.

TABLE II

STATISTICALLY SIGNIFICANT DIVERGENCES BETWEEN CURVES SHOWN IN FIG 1

The First Column indicates the given Pair of Curves under Comparison, and the Second, the Points along the Abscissa at which the Divergence is Significant at the $< .01$ Level.

Curves compared	Points of divergence significant at the .01 level*
Child and adult, A (length constant)	.25, .65, .75
Child and adult, B (area constant)	.45, .65
Child, A (length constant) and child, B (area constant)	.35, .65, .75
Adult, A (length constant) and adult, B (area constant)	.25

* The significances are only approximations, as the standard errors of the medians on which they are based were computed from ranked data.

2. The two B curves (constant area) show the same trends as the corresponding A curves, but in strikingly less degree, i.e., they are much flatter and closer together.

3. The divergence of the B curves from the corresponding A curves is such as to suggest that the size factor operates differentially in the two age groups; for any given width-length ratio the school children tend to prefer the larger rectangle, while the college students incline toward the more medium-sized one. (The Series B rectangles are the larger up through width-length ratio .45, and the Series A, thereafter.)

Summary and Conclusions

Two groups of *Ss*—100 children and 100 adults (women college students)—ranked 12 cardboard rectangles on the basis of aesthetic preference. The rectangles, which ranged in width-length ratio from .25 to .75, by steps of .10, comprised two series, one constant in length ($2\frac{3}{8}$ in.), and the other, constant in area (2.82 sq. in.). Curves for the two groups and two series, plotted in terms of median preference-rank, showed the following:

1. The curves for the constant-length series were very similar to those reported by Thompson. Except for an initial dip, the curve for the children showed a steady increase in preference for the wider rectangles, while that for the adults showed an increase only up to width-length ratio .65 (approximately the 'golden section'), and then a marked decrease.

2. The curves for the constant-area series showed the same general trends as those for the constant length, but to a strikingly less extent.

3. The divergence between the several curves was such as to suggest a differential effect of the size factor, with the children inclining toward the larger, and the adults, toward the more medium-sized rectangle for any given width-length ratio.

(Manuscript received September 4, 1946)

Reference

1. Thompson, G. G. The effect of chronological age on aesthetic preferences for rectangles of different proportions. *J. exp. Psychol.*, 1946, **36**, 50–58.

ACCURACY OF BINAURAL LOUDNESS MATCHING WITH REPEATED SHORT TONES *

BY W. R. GARNER

The Johns Hopkins University

Introduction

In most of the studies concerned with differential sensitivity to intensity and tonal duration, the procedure is to compare a short tone with a tone of somewhat longer duration. Such comparisons invariably are made with monaural listening, or binaural listening with the same stimuli in each ear, i.e., the stimulation is not *dichotic*. Usually when binaural differential sensitivity to intensity is studied, the work is aimed at the problem of localization of sound sources. Upton (**6**) and Ford (**1**) report measures on the binaural intensity-disparity threshold, but no studies have been done in which the binaural intensity-disparity threshold is measured as a function of tone duration. Furthermore, no studies are reported in which short tones are compared with short tones, even monaurally.

There are many times, however, when it is desirable to know how differential sensitivity varies as a function of duration when both the standard and the comparison tone are short. One of the areas of research in which such information is relevant is that mentioned above —the problem of localization of sound sources.

Control of time parameter.—The parameter of time in audition (or in any of the other senses) can be varied in two ways. One means of varying time is to change the duration of single stimuli, either repeated or in isolation. The other available means is to change the rate at which short stimuli are repeated. Either of these methods has the effect of increasing the total acoustic energy in a given period of time, although the two means may produce entirely different effects. These differences between methods may, however, provide us with valuable information concerning the nature of the auditory process.

Problem.—The experiments reported here are designed primarily to show how the duration and repetition rate of tones affect the binaural intensity-disparity threshold. Some investigation is also

* This research was carried out under Contract N5ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This article is Report No. 166-I-16 under that contract.

made of the effects of frequency and intensity on differential sensitivity. *O*s are given the task of matching tones in the two ears in loudness, when the tones in the ears differ only in intensity. The standard deviation of these matched values is used as a measure of differential sensitivity. With this technique, data on loudness relations of the two ears and the magnitude of differences between *O*s as compared with differences within *O*s are obtained.

Method

Apparatus

A schematic bloack diagram of the audio circuits used in these experiments appears in Fig. 1. A Hewlitt-Packard oscillator produces the sine-wave tone, and its output is fed into an electronic

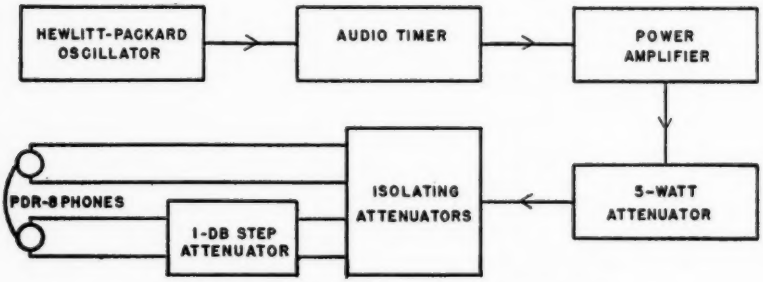

Fig. 1. A schematic block diagram of the apparatus

In addition to the equipment indicated, a Ballantine VTVM, a Dumont oscillograph, and a General Radio Type 713B oscillator were used for calibrations.

timer. The output of the timer is amplified and attenuated before being separated into two independently controlled channels by means of the isolating attenuators. The isolating attenuators are essentially mixing attenuators used in reverse, and serve both to isolate and to provide separate attenuation controls for the two channels. The two channels are led separately to each of the two earphones of a single headset. One of the channels contains a separate means of attenuation which the *O* uses to obtain the loudness match.

Audio timer.—The electronic timer used for controlling the repetition rate and duration of the tones has been previously described (4). The timer provides tone durations of from 1 to 2000 msec., and repetition rates from once every half-min. to 100 times per sec. A synchronizing mechanism allows each successive tone to be started at any part of the tone cycle, and, in these experiments, the tones always started at the time axis.

Earphones.—Acoustic energy was produced by two PDR-8 earphones which had been calibrated, when operating into a six cc. coupler, by the Permaflux Company. The phones were selected on the basis that they were the most nearly matched, with respect to frequency response, of 10 phones available. Both earphones were mounted in doughnut cushions.

Calibrations.—A Ballantine voltmeter, a Dumont oscillograph, and a General Radio 713B oscillator were used for calibration. All intensities were calibrated with a voltage sufficient to produce a sound pressure of 110 db *re* 0.0002 dynes/cm.^{2}, according to the calibration curves provided with the earphones. All sound pressures below that level were obtained by attenuating this signal.

Short tone durations were calibrated by counting the number of cycles of a 1000 cps tone on the oscillograph. Durations over 50 msec. were calibrated with a Standard Electric Clock, Model S-1. This same clock was used to calibrate repetition rates below five per sec. Repetition

rates greater than that were determined with the aid of Lissajous figures on the oscillograph. The low frequency output of the GR oscillator provided the reference frequency for this calibration.

Listening conditions.—All of the apparatus with the exception of the earphones and the one-db step attenuator were located in a small control room adjacent to a sound-deadened room in which observations are made. The observers could not see what conditions were being set in the equipment, although voice communication between the two rooms was possible by means of an intercom system.

Procedure

Observers.—In the majority of the experiments, six men (ages 18 to 25) acted as *O*s. These men were all well trained in auditory observation, each of them having served previously in similar experiments for a period of five months. During one phase of the experiments, 20 naive *O*s were desired, and these were secured from around the laboratory. Only one or two of these *O*s had had any previous experience with auditory observation.

Observations.—When observations were being made, *O* was seated in the sound-deadened room and told to adjust the headset so that it felt most comfortable to him. He was told that the same tone would appear in each earphone and that his task was to adjust the intensity of the tone in the right ear until both tones sounded equally loud. The one-db step attenuator used for this purpose had a total range of 45-db attenuation. When he had made his observation, he called the attenuation number to *E* in the outside room, and the number was recorded. *O* then removed the headset from his head while *E* made the necessary adjustments for the new conditions, at which time *O* replaced his headset and made his next observation.

The *O*s were required to remove the headset between observations to avoid the possibility that they could become adapted to a difference in intensity between the two ears. A total of either five or 10 observations was made for each condition, although no two observations for the same condition were repeated until all other conditions in a group had been used at least once. An *O* was not required to make continuous observations for more than an hour.

The intensities at the two ears were always adjusted so that the *O* could make the tone in the right ear obviously louder or weaker than the tone in the left ear. In between observations, however, the intensity in the right ear (the subject-controlled intensity) was adjusted by *E* in the control room. The amount of attenuation introduced by *E* varied over a range of 20 db, so that *O* could not form a position habit in his attenuation adjustments. The attenuation required for *O* to make the two tones *physically* equal in intensity was never the same for two successive judgments.

Measurements.—The measure recorded is always an attenuation score. From the attenuation scores, two measures are obtained. The average of all scores for a particular condition gives an attenuation score for equal loudness in the two ears. Thus, if the average score is 45 db, and the attenuation in the left ear (constant intensity) is 50 db, then the tones sound equally loud when the right ear is five db more intense than the left ear.

Measures of *differential sensitivity* (the binaural intensity-disparity threshold) are obtained by using the standard deviation (SD) of the observed attenuation scores. Many times in this paper standard deviations are reported for a combination of conditions. In these cases, variances, rather than standard deviations, are averaged, and the reported measure is the square root of the average variance. In actually computing such scores, variances for total groups of scores are computed, and then the variance attributable to differences between *O*s and conditions is subtracted. The standard deviation is the square root of this residual variance. This measure is mathematically equivalent to the square root of the average of all individual variances, i.e., the variance of each set of five or 10 scores for each individual under a given set of conditions. Thus, the variability caused by differences between *O*s or conditions is never included in the reported measures of variability, unless specifically stated.

Results and Discussion

Intensity and Frequency

Conditions.—Before investigating more thoroughly the factors of tonal duration and repetition rate on the binaural intensity-disparity

threshold, a preliminary experiment was performed to discover what effect, if any, intensity and frequency have on the threshold. For this experiment, four intensities, three frequencies, and three temporal conditions were chosen, making a total of 36 conditions. The intensities were 40, 60, 80 and 100 db *re* 0.0002 dynes/cm.^{2} The three frequencies were 250, 1000 and 4000 cps, and the temporal conditions were steady tones, and tones repeated 10 and 25 times per sec. When repeated tones were used, the tonal duration was 20 msec. Each of four *O*s made five loudness matches for each condition.

Effect of intensity.—The effect of intensity on the binaural intensity-disparity threshold is shown in Fig. 2. Each curve plotted in

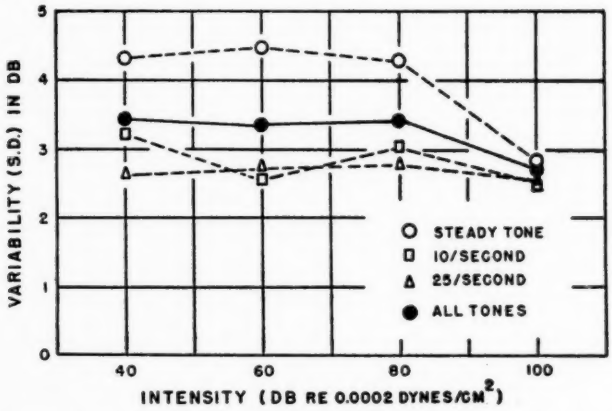

FIG. 2. The relation between intensity and variability of loudness matching as a measure of the binaural intensity-disparity threshold

Each plotted point for the dashed curves represents data combined from three frequencies, a total of 60 observations per point. Duration of the repeated tones is 20 msec.

that figure includes data from all three frequencies, and each curve is for a different temporal condition. The solid curve represents the average variability for all conditions. Intensity has little effect on variability of judgments over the range of the three lower intensities, although variability seems to decrease somewhat for the higher intensity. What is even more apparent from these curves, however, is the fact that variability is less for all conditions of interrupted tone than for steady tones. Differences between the two interrupted-tone conditions are less, but suggest that variability is slightly less for the faster repetition rate.

Effect of frequency.—Fig. 3 shows the effect of frequency on variability of judgments. There seems to be little difference between frequencies of 1000 and 4000 cps, although variability for the steady

tone is considerably less at 250 cps. Variability for repeated tones is affected very little by the frequency of the tones. It is not clear why variability should be less for the steady tone at 250 cps, since that result is not in conformity with data obtained with monaural listening. The difference between 250 cps and the other two frequencies, however, is statistically significant, since the probability of those differences occurring by chance is less than one in 200, as determined by the *F*-test.

Duration and Repetition Rate

Conditions.—On the basis of the preliminary results obtained, the factors of repetition rate and duration were investigated further, with

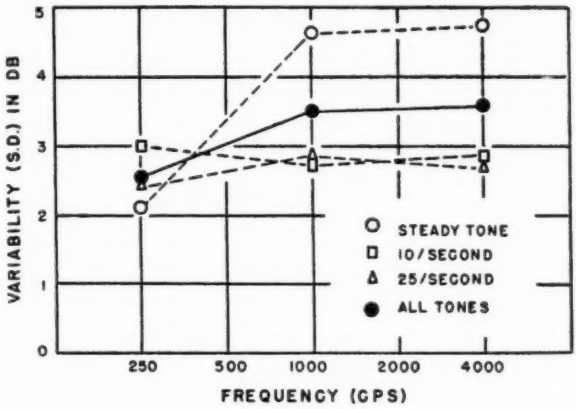

FIG. 3. The relation between frequency and variability of loudness matching as a measure of the binaural intensity-disparity threshold

Each plotted point for the dashed curves represents data combined from four intensity levels, a total of 80 observations per point. Duration of the repeated tones is 20 msec.

a 1000 cps tone at a 60 db intensity level. Variability of loudness matching was measured for six durations: 1, 10, 20, 50, 100, and 200, msec. Variability was also measured for a total of seven repetition rates: 1, 2, 5, 10, 25, 50, and 100 times per sec. Durations of 100 and 200 msec. were used at repetition rates of one per. sec. only. The one-msec. tone was used with all repetition rates, and the other durations with as many rates as is physically possible. A duration of 10 msec., for example, cannot be repeated 100 times per sec., since that would be a steady tone. Likewise, a duration of 50 msec. cannot be repeated 25 times a sec. For these measures, each of five observers made 10 loudness matches for each condition.

Effect of duration.—The effect of duration on the binaural intensity-disparity threshold as measured by the standard deviation is shown in Fig. 4. In this figure, data have not been included for repetition rates greater than 10 per sec., since so few durations were used at those rates. Durations of 100 and 200 msec. are for repetition rates of one per sec. only.

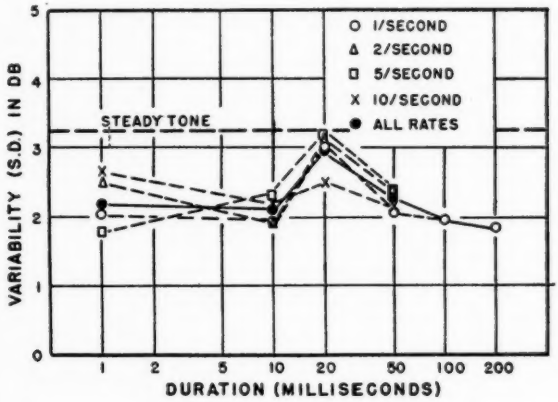

FIG. 4. The relation between tonal duration and variability of loudness matching as a measure of the binaural intensity-disparity threshold

Each dashed curve is for a different repetition rate as indicated. Durations of 100 and 200 msec. were used only at a rate of one/sec. The points for the dashed curves are based on 50 observations, and the solid curve represents the combined data. The standard deviation for steady tones is indicated by the horizontal dashed line.

The functions appearing in Fig. 4 are somewhat unusual. For durations of 20 msec. and larger, the sensitivity to changes in intensity improves as would be expected. All other results indicate that sensitivity of any kind improves with an increase in the duration. Below 20 msec., however, there is again an improvement in sensitivity, to an extent that the accuracy of loudness matching is almost as good with a one-msec. tone as with a 200-msec. tone. The shape of these curves suggests a discontinuity in the function, a break between 10 and 20 msec. This discontinuity may be real, since for durations under 20 msec. the stimulus is essentially a click. The stimulus is clearly a tone for longer durations, and it may be this difference which accounts for the unexpected nature of the functions.

Effect of repetition rate.—Fig. 5 shows the effect of repetition rate on variability. The two longest durations are not included in these data, since they are used only with a repetition rate of one per sec. While the curve for all durations suggests a slight increase in vari-

ability with higher repetition rates, still the differences are very slight, and it cannot be said that repetition rate has a serious effect on the binaural differences-threshold for intensity.

These results concerning the effects of both duration and repetition rate on variability, or difference-threshold, are somewhat surprising in view of the much greater effect the temporal factor has on the intensity-difference threshold when measured monaurally. In fact, in every other case where duration has been a variable, sensitivity of any kind is seriously reduced with tones as short as those used in these experiments.

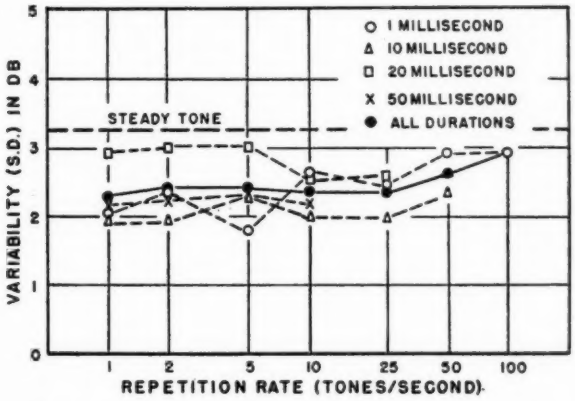

FIG. 5. The relation between repetition rate and variability of loudness matching as a measure of the binaural intensity-disparity threshold

Each dashed curve is for a different tonal duration as indicated. The faster rates do not involve the longer durations. The points for the dashed curves are based on 50 observations, and the solid curve represents the combined data. The standard deviation for steady tones is indicated by the horizontal dashed line.

There are two possible reasons why a changed duration or repetition rate has so little effect on differential sensitivity measured under these conditions. The lack of effect may be due to the fact that short tones are compared with short tones instead of long tones. This possibility should be checked monaurally. These results may also be due to the fact that binaural comparisons were made. It is possible that the usual change in differential sensitivity does not occur under conditions of binaural listening. There is the further complication that only binaurally can the tones to be discriminated be heard simultaneously. This simultaneity of presentation could make the nature of the discrimination quite different from that usually found with monaural listening.

Differences between Observers

It was mentioned above that all the scores of variability presented in the results do not include variability due to differences between *O*s. Such differences occur, however, and they should not be ignored. In fact, the most interesting result from these experiments was the realization that differences between *O*s are considerably greater than differences within *O*s. That is to say, an *O* can repeat his own observations with considerably more accuracy than he can duplicate another *O*'s observations.

In all of the measures obtained, it was the *O*'s task to match two loudnesses, and the binaural intensity-disparity threshold was measured only indirectly by the use of a variability score. The average of all his observations for a given condition is the measure of the point of subjective equality for the loudness of the tones in each ear. A succession of such observations makes it possible for us to plot equal-loudness relations for the two ears; that is, we can plot the intensity at one ear necessary to give a loudness equal to that produced by a given intensity in the opposite ear. Such a series of equal-loudness curves has been plotted in Fig. 6. Since the data for these curves were obtained from the preliminary experiment, each plotted point is the average of five observations.

It is clear from Fig. 6 that however accurate an *O* may be in repeating his own observations, he is not very accurate in terms of a physical equality. The two earphones used did not differ by more than one db at any of the frequencies used, so that any inability of the *O* to match the two tones physically was not due to the character of the phones. Furthermore, the really important fact is not that the matches do not agree with physical equality, but rather that the various *O*s agree so poorly with each other. Disparities between *O*s are as great as 25 db in some cases.

The fact that these differences are real is clearly demonstrated by the consistency of the differences. These measures are only based on five observations, but they are, nevertheless, extremely consistent. The rank order between various *O*s rarely changes.

To some extent these differences can be accounted for in terms of threshold differences between the two ears. If the curves shown in Fig. 6 had been plotted with sensation level measurements, the differences would not have been nearly as great. They would not, however, have disappeared entirely. Sometimes the differences in loudness are greater than the differences in threshold, and in one case at least, the disagreement with physical equality becomes consistently greater with higher intensities.

Even though the differences could be explained entirely on the basis of differences in threshold, the essential problem created by these

differences has not changed. The fact still remains that, at high intensities, equal loudness at the two ears is in very poor agreement with equal physical intensities.

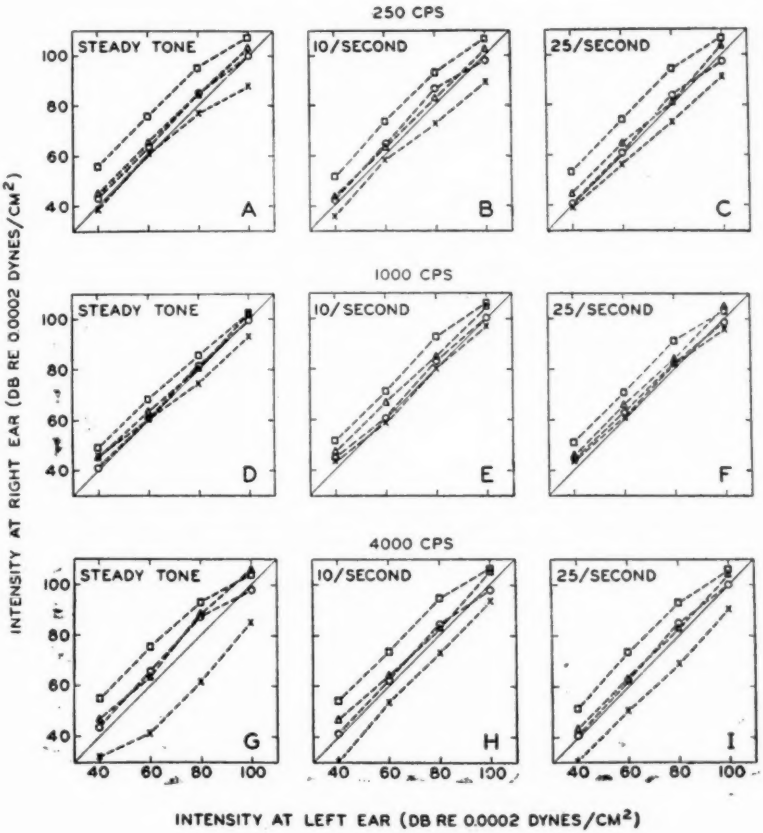

FIG. 6. Binaural loudness relations

These figures show the relative intensities at the two ears necessary to produce equal loudnesses. Each symbol is for a different *O*. The duration of the repeated tones is 20 msec.

It is, of course, possible to test for the significance of these differences between *O*s by means of the *F*-test, using variability within *O*s as one estimate of population variability, and variability between means as another estimate of population variability. Such variance ratios, when computed, are considerably larger than those necessary to be significant at the one percent level.

The possibility always exists that the five different *O*s used in these experiments were a poor sample, and that actually such large differences do not occur in the normal population. In order to secure a better estimate of population variability, observations were obtained on a total of 20 naive *O*s. These 20 *O*s were given the same task as the experienced *O*s, with a 1000 cps tone, and an intensity of 60 db *re* 0.0002 dynes/cm.^{2} Three different tones were used—a steady tone, and a repeated tone of 20 msec. with repetition rates of 10 and 25 per sec. For each of the three tones, five observations were obtained per *O*.

TABLE I

Two estimates of the variance of means scores are computed to test for the significance of differences between mean scores for equal loudness. Data are obtained from 20 naive *O*s.

	1000 cps tone		
	Steady	10/sec.	25/sec.
V_M	25.83	22.00	22.91
est. V_M	2.05	1.94	3.77
V_M/*est.* V_M	12.60	11.34	6.08
P	$\ll$.005	$\ll$.005	$\ll$.005

Table I shows the results of these observations. For each of the three tones, two estimates of the population variance of the means of sets of five observations are obtained. One of these estimates is the computed variance of the mean scores (shown as V_M in the table), and the other variance estimate is computed from the average variance of individuals (shown as *est.* V_M in the table). Small sample techniques are used in the computation. For all three tones, the probability that such differences between variance estimates could occur by chance is infinitesimal. Thus, we are forced to the conclusion that two tones of the same intensity are not of the same loudness if they are in two ears of the same individual. These differences are large and real.

Further evidence that these large differences between *O*s for the point of equal loudness are consistent is obtanied from the intercorrelations. The three product-moment correlations between the three different tonal conditions are .66, .72, and .71. These intercorrelations are surprisingly good when it is considered that the mean scores are based on only five observations.

The Problem of Sound Localization

These large differences between the loudnesses of the two ears are particularly interesting in respect to the problem of the localization of

sound sources. Stevens and Newman (5) have suggested that intensity differences between the two ears are primarily responsible for our ability to localize sound sources with frequencies above about 1000-2000 cps. Recent evidence by Ford (1) has indicated that intensity differences alone are a poor cue to localization even at the higher frequencies, for when *O*s have only intensity to facilitate their localization, they cannot achieve the same accuracy as they can in the open field condition. The data of Ford were obtained in a situation where the only task of the *O* was to turn his head toward the source of sound. This task should presumably give much better localization than the condition which Stevens and Newman provided, whereby the *O* was required to indicate the direction of the sound.

The data of the present experiments suggest the same conclusion that Ford reached, namely, that intensity alone provides a very inaccurate cue for the localization of sound sources. In the present experiments, the *O*s were not asked to localize a sound; their only task was to equate tones in loudness. Thus on the basis of the difference in task set, it is possible that the results obtained in localization experiments are not in conflict with these results. There remains, however, a rather difficult problem.

Measurements by Wiener (7) recently have shown that the maxmum intensity ratio between the two ears which occurs as a result of lateral displacement of a sound source of 4000 cps is 14 db. Yet in the present experiments, with the same frequency, *O*s differed by more than 20 db in determining the point of equal loudness. There are two possible explanations of these discrepancies.

In the first place, it is possible that such disparities in loudness would not occur in the open-field situation. These differences between the two ears may be due to such things as the way the phones fit on the head, or partial shielding effects of the outer ear. The fact that *O*s were required to remove the phones between each observation makes it improbable that there is any consistent effect due to a poor placement of the phones. On the other hand, variability was considerably greater at 4000 cps than at 250 cps, which suggests that some shielding effect might be taking place. If such an effect were present, it would presumably be greater at the high frequencies than at the low frequencies.

The other possible explanation of these discrepancies is in terms of learning. Even though there are differences in the loudness of the two ears, the *O* may learn unconsciously to compensate for these differences through experience, and thus is able to make a correct localization of a sound source even though loudness is not equal at the two ears.

The Problem of Methodology

In the experiments reported by Upton (**6**) on binaural difference thresholds for intensity, he reports no differences between the loudness at the two ears, although data for only one *O* are presented. Ford, however, using the technique which required a localization, not a loudness match, reports differences in average localization between his four *O*s equivalent to 3.5 db. These disparities are not nearly as great as those reported here.

The differences in methodology may in part, at least, account for the fact that large discrepancies are reported in this paper, but not in the previous papers. In Ford's experiments, the *O* could turn his head to one side and the other, and thus hear the maximum intensity ratio in each direction. Since his only task was to point his head toward the apparent source of sound, he could do so by making a judgment of half-way between the two extremes. Any inherent differences in loudness would thus tend to be obscured, since the Os were not actually making a loudness match.

In Upton's experiments, a method of limits was used in which the two tones started as physically equal. A discrepancy was introduced and increased until the *O* reported a difference. This method, then, would likewise obscure differences in the loudness of the two ears, since *O*s can detect the change without the tones being equal in the first place.

Two recent papers are of interest in respect to this problem of methodology. Koester and Schoenfeld (**2**) show that psychophysical judgments are made to the total context of stimuli, not just to the intended standard stimulus. In the case of pitch, at least, all judgments are shifted toward the mean level of all stimuli, and if more than one standard stimulus is used, then there will be constant errors in the judgments, directed toward the mean frequency. In a method of constants, all comparison stimuli become part of the context, and thus the point of subjective equality is forced toward physical equality if an equal number of stimuli are above and below the standard. Likewise, with a method of limits, the total context of stimuli is at the point of physical equality, and judgments will tend toward that mean value.

Postman (**3**) recently presented evidence that the constant error for loudness is greater when a method of average error is used than when a method of constant stimuli is used. He states that since an *O* is not forced to attend to the standard stimulus in a method of average error, his judgments are not as accurate. Postman recommends the use of a method of constant stimuli. In Postman's experiments, however, the fact that the constant errors with the method

of average error were very consistent is the best possible evidence that *O*s do attend to the standard stimulus. And since all comparison stimuli are part of the context of stimuli, constant errors will be minimized with a method of constant stimuli.

Contrary to the conclusion reached by Postman, it seems to the present author that the method of average error is preferable when the measurement being made is one of subjective equality. In other words, the best way to obtain a measure of equality is to ask the *O* to judge equality, not differences. The method of average error asks essentially that of the *O*. Both the methods of constant stimuli and limits, however, ask the *O* to judge only differences, or more accurately, changes. Differences in the point of subjective equality will be obscured with any method in which the *O* has no freedom in determining his total context of stimuli.

Summary and Conclusions

These experiments are concerned with the problem of binaural loudness matching. Measures of equal loudness for the two ears have been obtained, and the standard deviations of the loudness matches are used as a measure of the binaural intensity-disparity threshold. The following results and conclusions concerning the effect of various factors were obtained.

1. Variability of matching repeated short tones is unaffected by frequencies over the range 250 to 4000 cps. Steady tones, however, are matched more accurately at 250 cps.

2. Variability of matching both steady and interrupted tones is less with an intensity of 100 db *re* 0.0002 dynes/cm.^{2} than with lower intensities, but the effect of intensity is, on the whole, very small.

3. Variability of loudness matching is greatest with durations of 20 msec., and is smaller with either longer or shorter durations. Differential sensitivity is almost as good with one-msec. tones as with 200-msec. tones.

4. The repetition rate of short tones has no serious effect on the binaural intensity-disparity threshold.

5. Variability of loudness matching is greater with steady tones than with any form of repeated short tone, except in the case of a steady 250-cps tone.

6. Variability between *O*s in determining the point of equal loudness is much greater than the average variability within *O*s. *O*s differ in the intensity required for equal loudness at the two ears by as much as 25 db under certain conditions.

7. The fact that the inter-subject differences are much larger than any previously reported may be due to the methodology. It is sug-

gested that the method of average error is superior to other methods when measures of subjective equality are desired.

8. Binaural intensity disparity at best provides a very inaccurate indication of the displacement of a sound source. This inaccuracy is due primarily to the fact that equal loudness at the two ears corresponds very poorly to physical equality. The effect of various factors on the binaural intensity-disparity threshold is very slight compared to the inherent inaccuracy in matching loudness.

Acknowledgments

Mrs. Mary Hall and Mrs. Katherine Hill aided considerably in recording data and performing various calculations. Their assistance is gratefully acknowledged.

(Manuscript received for immediate publication June 19, 1947)

References

1. Ford, A. Dynamic auditory localization: 1. The binaural intensity disparity limen. *J. acoust. Soc. Amer.*, 1942, **13**, 367–372.
2. Koester, T., & Schoenfeld, W. N. The effect of context upon judgments of pitch differences. *J. exp. Psychol.*, 1946, **36**, 417–430.
3. Postman, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.*, 1947, **60**, 101–108.
4. Roush, R. Tone burst generator. *Electronics*, 1947, **20**, No. 7, 92–96.
5. Stevens, S. S., & Newman, E. B. The localization of actual sources of sound. *Amer. J. Psychol.*, 1936, **48**, 297–306.
6. Upton, M. Differential sensitivity in sound localization. *Proc. nat. Acad. Sci.*, 1936, **22**, 409–412.
7. Wiener, F. M. On the diffraction of a progressive sound wave by the human head. *J. acoust. Soc. Amer.*, 1947, **19**, 143–146.

PITCH CHARACTERISTICS OF SHORT TONES. I. TWO KINDS OF PITCH THRESHOLD *

BY J. M. DOUGHTY AND W. R. GARNER

The Johns Hopkins University

INTRODUCTION

Pitch in short tones.—Several changes in pitch occur when the duration of a tone is increased from some short time on the order of two or three msec. to one-half sec., and the three principal stages of of change have been described by Bürck, Kotowski, and Lichte (1). In the first stage, the *S* hears only a click having little or no pitch. In the second stage, the sensation is still predominantly a click, although some pitch character is apparent. In the third stage, pitch is easily identifiable and further increases in duration lead to little or no change in the pitch character of the tone. These authors state that the apparent pitch of tones falling into the second stage is different from the apparent pitch of tones falling into the third stage. The pitch of high tones is lowered by shortening duration, while the pitch of low tones is raised. Stevens and Davis (4), on the other hand, indicate that the pitch of tones falling into this second stage is always lower than the pitch of tones in the third stage, regardless of the frequency of the tone.

Purpose.—To study the relationship between pitch and the duration of a tone, then, becomes the purpose of this series of experiments. As a preliminary to further studies along this line, this particular experiment was designed to determine the limits of pitch change from the point at which pitch first becomes detectable to the point at which tone is present in the sound. These two points correspond to stages two and three noted above. The designation *click-pitch* threshold was assigned to the shortest duration at which the sound, predominantly a click, takes on a pitch quality which distinguishes it from another sound with a different central component. For instance, the click-pitch of a 125-cycle tone is lower than that of a 250-cycle tone. The designation *tone-pitch* threshold was assigned to the shortest duration at which the sound takes on a pitch quality about identical with the pitch of a tone of the same frequency but of a longer duration. In tone-pitch, tone, rather than click, dominates the sensation.

* This research was carried out under Contract N5ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This article is Report No. 166-I-14 under that contract.

Apparatus

A block diagram of the apparatus used in these experiments is shown in Fig. 1. A General Radio beat-frequency oscillator was used to generate the tone, and its output was fed into an audio timer, a specially designed electronic circuit which allowed the tone to 'come on' for variable durations. From the timer, the signal was led successively to an amplifier, a db-step attenuator, and an impedance matching transformer (LS-33). Two pairs of Permaflux earphones (Type PDR-8) serve as receivers, one pair for *O*, and another pair for *E*. These receivers were mounted in doughnut headsets (NAF-48490). Connected in parallel with the earphones for purposes of calibration were a Ballantine Voltmeter and a Dumont (Type 208-B) cathode-ray oscillograph.

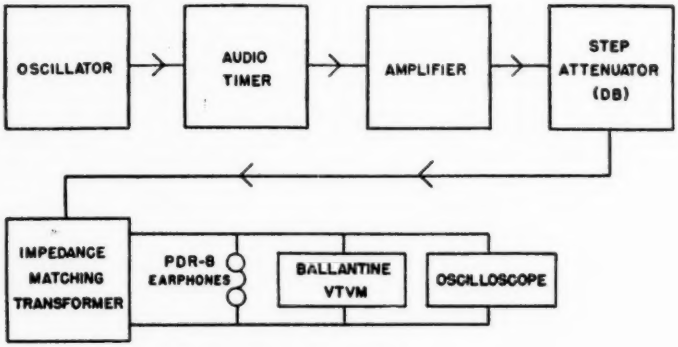

Fig. 1. Block diagram of the apparatus

Earphone calibration.—The earphones used in these experiments had been calibrated by the Permaflux Co. for operation into a 6-cc. coupler, and all phones had a response which was reasonably flat between 100 and 8000 cps. For all calibrations of reference intensity, the voltage across the phones was set at a level approximately equal to 110 db *re* 0.0002 dynes/cm.^{2}, as determined by the calibration curves. That is to say, the reference voltage produced a sound pressure in a 6-cc. coupler equal to 110 db *re* a pressure of 0.0002 dynes/cm.^{2}

Audio timer.—The audio timer used in these experiments deserves a further brief description. This timer was designed and built by R. G. Roush of this laboratory (3). It consisted of an electronic circuit which provided a self-repeating tone of variable duration. The tone would repeat at rates ranging from once every 20 sec. to 100 times per sec. The duration of the tone could be varied from less than one msec. to over two sec. by means of a three position switch and a vernier control. Special controls were also available making it possible to start the tone at any part of its cycle or, alternatively, the tone could be allowed to start at a random point in its cycle. During most of this experiment, the timer was allowed to operate in the latter manner. Fig. 2 shows an oscillogram of the output of the timer, with two tones starting at different parts of the cycle.

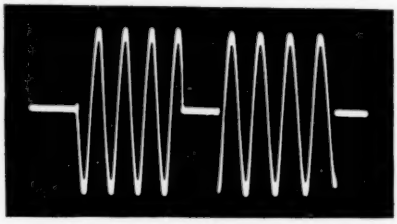

Fig. 2. An oscillographic picture of the output of the audio timer

Each of the two tones starts at a different part of its cycle. The first tone starts at the baseline (180 degrees from the beginning of the cycle), and the second tone starts approximately at its negative peak (270 degrees from the beginning of the cycle).

PROCEDURE

Selection of subjects.—In the preliminary stages of the experiment, six *S*s—all previously inexperienced in auditory research—were used. In a very short while it became apparent that the threshold judgments which were required of the *S*s were very difficult for a naive observer to make. While there was an overall general agreement between the *S*s, still some of the men had a great deal of difficulty in maintaining a constant criterion, which made it hard to determine the exact point of the thresholds. In order to obtain greater reliability for the measures, two of the more promising *S*s were selected from the original six; they were given rather intensive training, and all the data presented here were later obtained from them. The two *S*s were selected solely on the basis of reliability of observations.

Criteria for thresholds.—During this preliminary training, and during some exploratory investigation by the authors, two criteria for the thresholds were agreed upon.

1. The threshold for *click-pitch* was defined as the shortest duration of a tone which allowed the tone to have some pitch character to it. The sensation is still predominantly one of a click, but the click has sufficient pitch quality to enable the *S* to estimate its general frequency range.

2. The threshold for *tone-pitch* was defined as the shortest duration of a tone in which the tonal character, rather than the click character, was dominant. The general quality of the sensation should change very little with further increases in the duration of the tone.

In the case of click-pitch, the sensation is essentially instantaneous—there is no temporal aspect to the tone except that it seems very short. In the case of tone-pitch, the tone does have a temporal character—the beginning of the sensation is a click, followed by the tone, and then followed by another click when the tone ceases. The designations click-pitch and tone-pitch were chosen because in the first case the sensation is definitely one of a click, while in the latter case the sensation has become tonal.

Experimental plan.—The experiment was divided into two parts. The first part was concerned with the relation between the duration of tone required for the two thresholds and frequency of the tone. Seven frequencies were selected: 125, 250, 500, 1000, 2000, 4000, and 8000 cps, and for these frequencies the intensity level was kept constant at 110 db *re* 0.0002 dynes/cm.^{2} The second part was concerned with the relation between the duration of tone for the two thresholds and intensity of tone. Seven intensities were selected for these measures: 90, 80, 70, 60, 50, 40, and 30 db *re* 0.0002 dynes/cm.^{2} Three representative frequencies were used for the various intensity levels: 125, 1000, and 8000 cps.

Conditions of observation.—While the data for these various conditions were being obtained, one *S* and the *E* were located side-by-side in the same room with the apparatus, and the *S* was allowed to adjust the headset so that it fitted most comfortably on his head. He was seated in a classroom chair having an arm rest which provided support for his arm while he made the adjustments on the timer necessary to obtain his threshold. The *E* varied the frequency or the intensity in the presence of the *S* and in a random manner. For each condition set by the *E*, the *S* was asked to adjust the vernier dial which controlled tone duration until the duration gave a sensation corresponding with the criterion given. While the *S* was making his observation, the tone repeated automatically once per sec.

Judgments were restricted to just one type of threshold until all the frequencies (or the three frequencies at each of the selected intensity levels in the second part) had been presented. At this point, a short rest period of two min. was given. The judgments were then shifted to the other type of threshold and the process repeated.

Number of observations.—In the first part of the experiment, five judgments for each of the two types of threshold were obtained at each frequency in the course of the experimental period. In the second part, only two judgments were obtained for each threshold at each intensity level at each of the three frequencies, if these two judgments agreed within two msec. or less. Otherwise, five judgments were made. When about half of the total number of judgments to be made in the experimental period had been made, the *S* was given a 10 min. rest period. The average length of the experimental periods was approximately 1¼ hours. There were two such periods per week for each *S*.

In making his judgments, the *S* was instructed to ignore the scale markings on the duration vernier dial, either by closing or averting his eyes. This precaution was necessary to eliminate visual cues on the basis of which the *S* could conceivably make his judgments.

When the click-pitch or tone-pitch threshold had been obtained to his satisfaction, the *S* read off the vernier setting and the *E* recorded the reading. The five such vernier readings obtained for each type of threshold at each frequency in the first part, and the two or more readings obtained for each type of threshold at each intensity level for the three frequencies in the second part were averaged by the *E* at the end of the experimental period and converted into msec. values.

Computation of duration.—Previous calibrations of the vernier dial had shown the relation between dial setting and duration to be linear within the range used, so that arithmetic averaging of the vernier settings was justified. The actual measurement of duration was accomplished by increasing the repetition frequency of the tone until its rate could be synchronized with an oscillograph sweep without producing flicker. The vernier dial was then set at the average reading, and the number of cycles of a 1000 cps tone were counted. This method provides accuracy well within the limits of ¼ msec.

Results

Preceding the experiment proper, approximately 500 judgments, over a period of two weeks, were made by each of the two *S*s on these thresholds. The criteria for the thresholds were constantly emphasized during this training period.

Training periods.—The time to make a single judgment decreased steadily during the training period. In the early stages of training, an *S* would take as long as five min. to make a single judgment. In the final stages, the average length of time was cut to about 30 sec. Judgments at the low frequencies of 125 cps and 250 cps remained the most difficult to make throughout the entire experiment, and invariably took the longest to make. This was probably due to the phase differences in the onset of the sound—a crucial factor at low frequencies. Phase at the onset of the tone was not controlled during the running of the experiment, but a post-experimental investigation of its effect indicated that the effect was non-existent at frequencies of 250 cycles and above.

By the end of the training period, the *S*s reported that judgments of both click-pitch and tone-pitch were relatively easy to make at all frequencies higher than 250 cycles, with the click-pitch threshold being the more clear-cut of the two.

Preceding the second part of the experiment, another training period was given during which the *S*s made approximately 250 judgments for the two types of pitch throughout the intensity range stated and over the three frequencies used. This training was only for the purpose of giving the *S*s practice at recognizing click-pitch and tone-pitch at low intensity levels.

Results: Effect of Frequency

Table I shows the relation between frequency and duration of a tone in msec. for the click-pitch and tone-pitch thresholds. The thresholds for each frequency are based on the average of 40 observations obtained from the two *S*s over four experimental periods.

TABLE I

THE RELATION BETWEEN FREQUENCY AND DURATION OF A TONE IN MSEC. FOR TWO KINDS OF PITCH THRESHOLDS

The intensity is 110 db *re* 0.0002 dynes/cm.²

	Frequency (cps)						
	125	250	500	1000	2000	4000	8000
Click-pitch	17.9	10.5	5.8	4.4	4.6	4.1	4.0
Tone-pitch	24.2	16.8	12.9	10.2	10.2	8.8	9.6

For all observations the intensity level remained constant at 110 db *re* 0.0002 dynes/cm.²

Fig. 3 is a graphical representation of the data from Table I, and it can be seen from this figure that click-pitch and tone-pitch follow

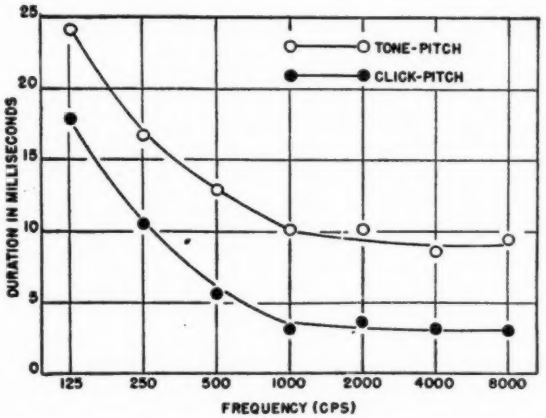

FIG. 3. The relation between frequency and the tonal duration required for two types of pitch The data for these two curves are obtained from Table I, and each plotted point represents the average of 40 observations obtained from two *S*s. The intensity of the tone is 110 db *re* 0.0002 dynes/cm.²

two very distinct curves throughout the frequency range from 125 cps to 8000 cps. The duration threshold for the perception of *click-pitch* is shorter at all frequencies than that for perception of *tone-pitch*, which is, of course, to be expected.

The thresholds for both types of pitch are longest at the low frequencies of 125 cps and 250 cps. They become shorter in the range of frequencies to which the ear is most sensitive, i.e., 1000 cps to 4000 cps, and remain fairly constant throughout this range of frequencies and up to 8000 cps.

Difference between thresholds.—Two facts are evident from these curves. In the first place, the difference in duration required for the click-pitch and the tone-pitch thresholds is not a relative matter, but rather is a constant. For example, it does not take twice as great a duration to go from click-pitch to tone-pitch regardless of the duration required for click-pitch, but it appears to require about six msec. more duration, for any frequency and any duration. The maximum difference in duration between any two thresholds is 7.1 msec., and the minimum difference is 4.7 msec. While there is some tendency for the time difference to become less as the frequency is increased (and as the duration for click-pitch is decreased), still that trend is very small compared to the relative constancy of the difference. Thus, it would seem that only increased duration, and not number of cycles, is responsible for the change from click-pitch to tone-pitch.

In the second place, as the frequency is changed, it requires neither a constant duration nor a constant number of cycles to obtain either of the thresholds. For frequencies above 1000 cps, duration is relatively constant. For frequencies below 500 cps, there is a tendency for the number of cycles to remain constant (more true for click-pitch than tone-pitch), although this tendency is far from exact. In other words, as long as the minimum duration required for click-pitch involves at least two to three cycles, that duration will be constant. When the frequency is so low, however, that an insufficient number of cycles would be involved, the duration must be increased so that at least two complete cycles are encompassed in the tone. For example, for frequencies of 125, 250, and 500 cps, the duration is not constant, although in each case the duration is equivalent to between two and three complete cycles (click-pitch). For frequencies above that, however, the duration is constant, indicating that the number of complete cycles increases proportionate to the frequency.

Results: Effect of Intensity

Table II shows the relation between intensity level and duration of a tone in msec. for the click-pitch and tone-pitch thresholds. The data from this table are plotted in two graphs: Fig. 4 shows the relation between intensity and tonal duration required for a tone-pitch experience, and Fig. 5 shows the relation between intensity and tonal duration required for a click-pitch experience.

For each of the three selected frequencies—125 cps, 1000 cps, and 8000 cps—the intensity level was decreased to the point at which judgments of the two types of pitch became too difficult to make. Table II indicates that this point was at 70 db for 125 cps, 30 db for 1000 cps, and 50 db for 8000 cps. From Figs. 4 and 5, it is apparent that the duration thresholds for the perception of both tone-pitch and

TABLE II

THE RELATION BETWEEN INTENSITY AND THE DURATION OF A TONE IN MSEC. FOR TWO KINDS OF PITCH THRESHOLDS

Intensity in db *re* 0.0002 dynes/cm.²	Frequency (cps)					
	125		1000		8000	
	Click-pitch	Tone-pitch	Click-pitch	Tone-pitch	Click-pitch	Tone-pitch
110	17.9	24.2	4.4	10.2	4.0	9.6
90	15.7	26.2	6.5	15.6	5.8	16.4
80	17.8	30.8	7.0	16.4	6.2	17.9
70	20.8	39.9	6.6	16.4	5.8	18.8
60			7.1	17.6	6.5	19.8
50			7.5	18.9	8.8	21.2
40			8.4	20.7		
30			10.4	21.6		

click-pitch increase as the intensity level is decreased. For instance, the duration necessary for the perception of *tone-pitch* at 1000 cps at an intensity level of 110 db is 10.2 msec. At 80 db it is 16.4 msec. At 50 db it is 18.9 msec., and at 30 db it has increased to 21.6 msec. (see Table II). The duration necessary for the perception of *click-pitch* at 1000 cps at 110 db is 4.4 msec. When the intensity has been decreased to 30 db, the duration necessary for the perception of click-pitch has increased to 10.4 msec.

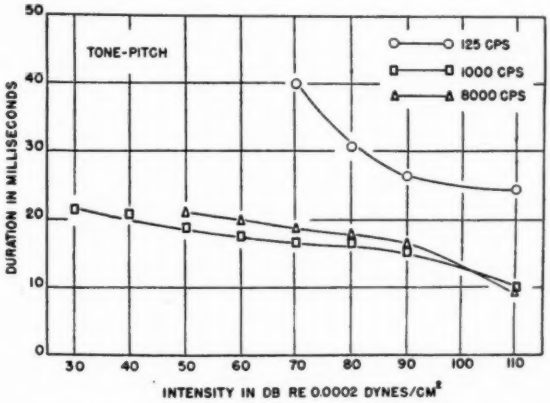

FIG. 4. The relation between intensity and tonal duration required for a tone-pitch experience The data for these curves are obtained from Table II, and each plotted point represents the average of 16 observations obtained from two *S*s.

Intensity thresholds.—This same trend is evident for both thresholds at 8000 cps, although at 125 cps the rise is somewhat greater for

tone-pitch. This is probably due to the fact that the 125 cps tone was much closer to the threshold than were the other tones, and it would be expected that the duration thresholds would increase more rapidly as they approached the threshold for intensity. There seems to be no explanation of the apparent reversal shown for click-pitch at 125 cps.

The two thresholds do not change at the same rate as the intensity is changed. The threshold for tone-pitch changes much rapidly than the threshold for click-pitch.

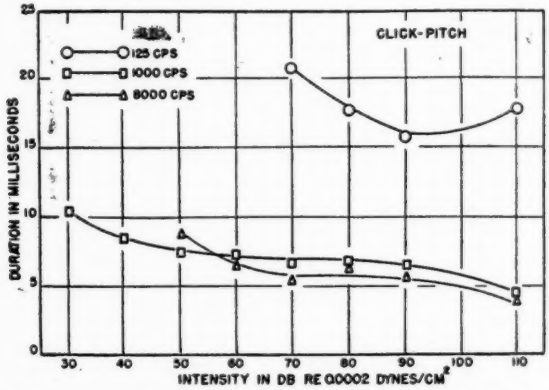

FIG. 5. The relation between intensity and tonal duration required for a click-pitch experience The data for these curves are obtained from Table II, and each plotted point represents the average of 16 observations obtained from two *S*s.

Interaction of frequency and intensity.—In Fig. 4, it can be seen that the two curves for 1000 and 8000 cps cross as the intensity becomes lower, and that at lower intensities, the duration threshold for 8000 cps becomes greater than that for 1000 cps. The separation of these two curves becomes consistently greater as the intensity is decreased, indicating that if the curves plotted in Fig. 3 had been obtained with a lower intensity, there would have been a rise in the duration at 8000 cps. This same crossing over does not occur with the click-pitch curves until the intensity is very low, and is probably due only to the fact that the 8000 cps tone was much closer to absolute threshold than the 1000 cps tone.

Comparison with other results.—In the paper by Bürck, Kotowski, and Lichte (1) previously cited, results were presented showing the relation between the duration threshold for pitch and frequency. Although the authors pointed out in that paper that pitch goes through three stages as the duration of a tone is increased, they did not state which of these three stages corresponded to the thresholds

they measured. The curves they present, however, show a rise in duration at the higher frequencies.

As pointed out above, if the curves relating tone-pitch thresholds to frequency (Fig. 3) had been obtained with a lower intensity, this same rise at the higher frequencies would have been obtained in these experiments. Although Bürck, Kotowski, and Lichte did not state the intensity at which they were working, it seems quite probable that it was lower than that used in these experiments for the comparable curves. Table III gives a comparison between their results

TABLE III

A Comparison between the Results Obtained by Bürck, Kotowski, and Lichte and the Results of This Experiment

	Minimum Durations Required for:	
	Tone	Tone-pitch
	Bürck, Kotowski, and Lichte	This experiment (80 db *re* 0.0002 dynes/cm.^{2})
125 cps.	32 msec.	30.8 msec.
1000 cps.	13 msec.	16.4 msec.
8000 cps.	19 msec.	17.9 msec.

and those obtained for tone-pitch in this experiment at 80 db *re* 0.0002 dynes/cm.^{2} From this comparison, it seems clear that they were obtaining measures for what has here been called tone-pitch, and that the two sets of data are not in serious disagreement. The greater rise at 8000 cps shown in their data is due primarily to a single S, and, considering the difficulty of obtaining these measures, the disagreement is not at all serious.

Discussion

Physical nature of short tones.—In order to understand fully the nature of the results presented here, it is necessary to have some understanding of the physical nature of tones of very short duration. An oscillation which is turned on for a very brief period of time does not consist of a pure tone, since frequencies other than that of the period of oscillation are produced when a tone is changed from zero to some finite amplitude of oscillation. A good many different frequencies, all adding up in the proper phase relation, are required to produce any oscillation of a given period of time, and as that period of time becomes smaller, more frequencies are required.

Distribution of energy.—By means of a Fourier analysis, it is possible to compute the distribution of energy in a very short tone. Two such spectral distributions of energy, for two different durations, are shown in Figs. 6 and 7, and in Fig. 7 the duration is half as long as that in Fig. 6. Several characteristics of such spectral distributions can be noted from an examination of these two figures. The maxi-

mum energy is always found at the frequency of oscillation, and the other energy is spread symmetrically on either side until no energy occurs at one particular frequency. The frequency of these first null points in the spectral distribution is defined by $f_0 \pm 1/d$, where f_0 is

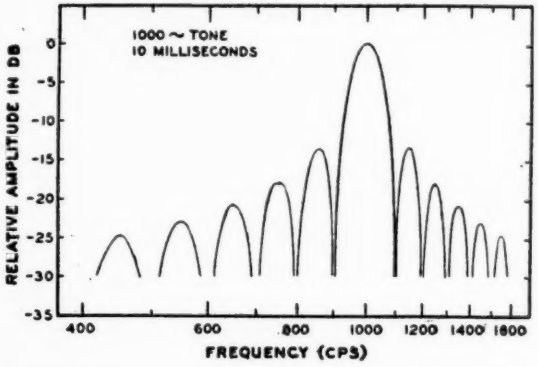

FIG. 6. The spectral distribution of the energy in a 1000 cps tone lasting for ten msec. There is additional energy at frequencies greater and less than those shown here, but it is very small compared to the energy in the main band.

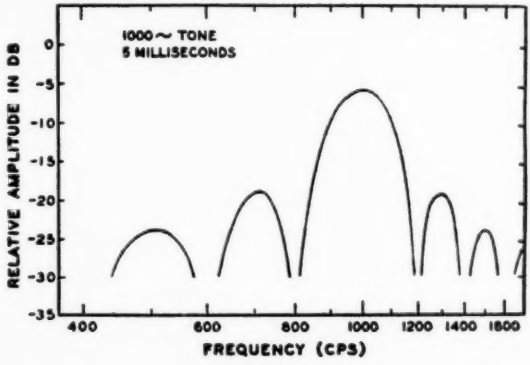

FIG. 7. The spectral distribution of the energy in a 1000 cps tone lasting for five msec. As in Fig. 6, there is a small amount of energy at frequencies above and below those shown here. The reference level is arbitrary, but is the same for Fig. 6 and Fig. 7.

the frequency of oscillation, and d is the duration of the tone. From that frequency on either side of the central frequency, the energy is distributed in bands of continuous frequencies and the width of these bands is defined by $1/d$, although the energy in these bands is quite small compared to the energy in the main band of frequencies.

It can be seen that the bandwidth of this main band of frequencies is defined by $2/d$, although $1/d$ is a good approximation of the bandwidth of the energy, if the half-power points are used to define the main bandwidth. Thus, if the duration of a tone is doubled, the bandwidth of frequencies involved in producing that tone is halved, and it can be said that the bandwidth of energy is inversely proportional to the duration of a tone. Considering this fact, one should expect that the longer the duration of a tone, the more it should sound like a tone rather than a click.

Relation to click-pitch.—It is probably this relationship to which the ear is responding for a click-pitch threshold. The duration of the tone must be increased until the bandwidth of energy is narrow enough that a reasonably well defined area of frequencies results. The ear responds as though it were stimulated with an instantaneous bandwidth of energy, and if that bandwidth of energy is narrow enough, some pitch character will result.

Temporal relations.—The physical nature of a short tone has not been adequately described, however, if only the spectral distribution of energy is considered. While it is true that all the frequencies shown in Fig. 7, for example, are necessary to describe a tone with an oscillation of 1000 cps for five msec., with a constant amplitude, still all those frequency components do not occur at the same time. As the tone continues, a greater and greater percentage of the energy is at the main frequency component, while at the beginning and end of the tone, the majority of the energy is at frequencies other than the central frequency. This extraneous energy occurs primarily where a tone is turned on and off instantaneously, but the longer the duration of the tone, the less there is of this energy. Thus in a longer tone what one actually hears is a click followed by a tone, with another click at the termination of the tone.

Tone-pitch.—It is probably this characteristic of short tones, as well as of the ear, which determines the threshold for tone-pitch. When the tone has been on long enough for the ear to distinguish the clicks at the beginning and end, and a tone can be heard in between the clicks, the tone-pitch threshold has been reached. In such a sensation, the tone is dominant, and can be distinguished from the clicks. In the click-pitch, the on- and off-click cannot be distinguished, and the ear responds as though the stimulation were instantaneous.

Thus, the click-pitch threshold is a question of frequency spectrum, while the tone-pitch threshold is a question of time relations in the ear. Considering these relations, we should not be surprised that the difference between the click-pitch and tone-pitch thresholds is a constant rather than a ratio.

Relation to other experiments.—This distinction seems justified in view of other work done by Bürck, Kotowski, and Lichte (**2**), who measured the difference in onset time required for two tones in order for the two tones to be heard as starting at different times. The times measured under these conditions were very similar to the times required for tone-pitch, and these authors postulate that the onsets of two tones can be distinguished when the clicks resulting from their onsets can be distinguished temporally. In other words, when the clicks merge, no difference in onset time can be heard.

Intensity relations.—These relations also help understand why the two curves relating thresholds for click-pitch and tone-pitch to intensity have different slopes (Figs. 4 and 5). In the case of click-pitch, where the change with intensity is relatively slight, the only factor determining the threshold is the relative narrowness of the band of energy, and this does not change as a function of intensity. Only when the intensity of the tone approaches the threshold of hearing does there seem to be any serious change in the duration required for click-pitch.

On the other hand, the slight tone which one hears between two clicks in the case of the tone-pitch threshold may not be nearly so loud as the click, and is in fact much nearer the threshold than the click. Thus, an increased duration becomes necessary to compensate for the loss of intensity, in order for the tone to be heard at all.

Tone-pitch and pitch discrimination.—Another study, reported by Turnbull (**5**), is relevant to this discussion. Turnbull was concerned with the relation between pitch discrimination and tonal duration. He reports that for a 1024 cycle tone effective pitch discrimination breaks down when the duration of the tone is shortened beyond approximately 17.3 msec. This duration is comparable to the tone-pitch threshold of 16.4 msec. reported in Table II for a 1000 cycle tone at an intensity level of 70 db. (This intensity level is about equivalent to the level of the Turnbull study, which was 60 db above threshold.) Though it is possible to make some pitch discrimination at shorter durations, the task is very difficult, and at durations of less than 10 msec., pitch discrimination at 1000 cps is effectively nil.

The figures reported by Turnbull as minimum durations for effective pitch discrimination at 125 cycles and 8129 cycles are not in such close agreement with our figures for tone-pitch at 125 cycles and 8000 cycles at the lower intensity levels as in the case of 1024 cycles, although both show longer durations necessary at these frequencies. Nonetheless, the agreement at frequencies in the vicinity of 1000 cycles does seem to indicate that *the tone-pitch threshold of this experiment defines the limit at which pitch is distinctive, and, as a consequence, the limit for effective pitch discrimination.* Though pitch is

present down to durations closely approximating our click-pitch thresholds, there is not enough to provide a basis for pitch discrimination.

Variability of observations.—The thresholds for tone-pitch and click-pitch in the first and second parts are by no means exact measures. The judgments are very difficult to make when the *S*s are relatively inexperienced. As indicated earlier, the two *S*s selected for this experiment were part of an original group of six. Variability was so great in this original group that it was thought best to work with only two of the more promising *S*s. Even during the running of the experiment, these two *S*s showed much variation within one experimental period, from day to day, and from *S* to *S*. For 125 cps, variations ran as high as nine msec. during one experimental period. At higher frequencies the variations were less but, even here, they were on the order of two or three msec. in one experimental period. Illustrating the differences between *S*'s, the average of one *S*'s judgments for tone-pitch at 125 cycles over the four experimental periods was 15.9 msec. and the average of the other *S*'s judgments was 19.9 msec. Tone-pitch judgments at 1000 cycles for one *S* average 9.8 msec.; for the other, 10.6 msec. At 8000 cycles the two averages were 8.8 msec. and 10.5 msec.

Besides the difficulty of the judgments, other factors may account for the variability, especially at the lower frequencies. Adaptation and phase at the start of the tone appear to be two such factors.

Adaptation.—The effect of adaptation was studied after the experiment had been completed. Readings for both types of pitch threshold taken over a 45-min. period, at a frequency of 125 cps, show an average shift of about three msec. in the thresholds. This variation tends to be less when changes are made in the stimulus to which the *S* is attending, i.e., when frequency or intensity is altered regularly.

Starting of tone.—At 125 cps, the part of the cycle at which the tone starts definitely affects the thresholds for both tone-pitch and click-pitch. In an investigation of this effect (carried out after completing those above) the average of 14 observations for each *S* showed that the duration necessary for the perception of click-pitch increased 3.5 msec. when the tone was turned on at the peak of the cycle instead of at the zero axis. An increase of about four msec. was noted for the tone-pitch threshold for the same conditions. Phase at the start of the tone apparently has no effect at the other frequencies used in the experiment, i.e., from 250 to 8000 cps.

Conclusions

1. The click-pitch experience, as defined in this paper, is an experience of instantaneous stimulation of the ear by a bandwith of

frequencies, and, as such, probably defines the widest bandwidth of energies which may be used to stimulate the ear when some identification of pitch is required.

2. The tone-pitch experience is, on the other hand, an experience primarily of two clicks, one at the beginning and the other at the end of the tone, with a tonal sensation between the two clicks. Tone-pitch thresholds are determined primarily by temporal conditions, i.e., by the ability of the ear to distinguish in time the two clicks occurring when a tone is turned on and off.

3. The duration thresholds for tone-pitch probably define the lower limit of duration of a tone which may be used in order for the pitch character of the tone to remain relatively constant, and in order to have reasonable pitch discrimination.

Summary

Problem.—As the duration of a tone is increased from sonething on the order of two msec. to approximately one-half sec., the pitch of the tone goes through three stages. In the first stage, only a click is heard. In the second stage, the sensation is still that of a click, although the click has some pitch character to it. In the third stage, the tone definitely has a pitch character, and that character will change little, if at all, with further increase in the duration of the tone.

Results.—In the experiment reported here, duration-thresholds have been obtained for these latter two stages, called *click-pitch* and *tone-pitch*, respectively. It was found that these thresholds vary to some extent with both frequency and intensity, and that the difference in duration between the two thresholds is relatively constant for all conditions. The duration required for both types of threshold decreases as the frequency is increased to about 500 cps. Beyond that the threshold is relatively constant. Likewise, a decrease in intensity results in an increase in duration for both thresholds. These results agree fairly well with results previously published by other authors, which are somewhat less complete.

The meaning of these results is explained in some detail in reference to the spectral distribution of energy of short tones, and also in relation to the time patterns of the frequencies involved in short tones.

Conclusions.—It is concluded that the thresholds for *tone-pitch* probably define the lower limits of duration which may be used if the pitch character of the tone is to remain relatively stable, and if reasonably accurate pitch discrimination is required. With shorter durations, pitch discrimination breaks down rapidly.

(Manuscript received for immediate publication June 13, 1947)

REFERENCES

1. BÜRCK, W., KOTOWSKI, P., & LICHTE, H. Der Aufbau des Tonhöhenbewusstseins. *Elek. Nachr.-Techn.*, 1935, **12**, 326–333.
2. BÜRCK, W., KOTOWSKI, P., & LICHTE, H. Die Hörbarkeit von Laufzeitdifferenzen. *Elek. Nachr.-Techn.*, 1935, **12**, 355–362.
3. ROUSH, R. G. Tone burst generator. *Electronics*, 1947, **20**, No. 7, 92–96.
4. STEVENS, S. S., & DAVIS, H. *Hearing: Its Psychology and Physiology*. New York: John Wiley, 1938.
5. TURNBULL, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.

ERRATA

In the article by J. C. Franklin and J. Brozek, this JOURNAL, 1947, **37**, 16–24, the following errors occurred:

1. The denominator of the formula for V_{wG} should read $k(n - 1)$ instead of $n(k - 1)$ on p. 22. No errors are involved in the published computations.

2. Although the titles of Figs. 1 and 3 are correct, the two graphs were interchanged.